# HIGH THROUGHPUT SEQUENCING-BASED TRANSCRIPTOME ANALYSIS OF DIFFUSE LARGE-B-CELL LYMPHOMA PATIENTS WITH SAMPLES TAKEN AT DIAGNOSIS AND AFTER THERAPY RELAPSE: A FEASIBILITY STUDY TOWARD DEVELOPING PERSONALIZED THERAPIES

by

Maryam Abdul Ahad

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science (MSc) in Biology

The Faculty of Graduate Studies
Laurentian University
Sudbury, Ontario, Canada

## THESIS DEFENCE COMMITTEE/COMITÉ DE SOUTENANCE DE THÈSE
**Laurentian Université/Université Laurentienne**
Faculty of Graduate Studies/Faculté des études supérieures

| | |
|---|---|
| Title of Thesis<br>Titre de la thèse | HIGH THROUGHPUT SEQUENCING-BASED TRANSCRIPTOME ANALYSIS OF DIFFUSE LARGE-B-CELL LYMPHOMA PATIENTS WITH SAMPLES TAKEN AT DIAGNOSIS AND AFTER THERAPY RELAPSE: A FEASIBILITY STUDY TOWARD DEVELOPING PERSONALIZED THERAPIES |
| Name of Candidate<br>Nom du candidat | Abdul Ahad, Maryam |
| Degree<br>Diplôme | Master of Science |
| Department/Program<br>Département/Programme | Biology |
| Date of Defence<br>Date de la soutenance | October 27, 2017 |

**APPROVED/APPROUVÉ**

Thesis Examiners/Examinateurs de thèse:

Dr. Hoyun Lee
(Co-Supervisor/Co-directeur de thèse)

Dr. Rebecca McClure
(Co-Supervisor/Co-directrice de thèse)

Dr. Kabwe Nkongolo
(Committee member/Membre du comité)

Dr. Ingeborg Zehbe
(External Examiner/Examinateur externe)

Approved for the Faculty of Graduate Studies
Approuvé pour la Faculté des études supérieures
Dr. David Lesbarrères
Monsieur David Lesbarrères
Dean, Faculty of Graduate Studies
Doyen, Faculté des études supérieures

**ACCESSIBILITY CLAUSE AND PERMISSION TO USE**

# ABSTRACT

Traditionally, prognostic information and selection of therapies for individuals with cancer is based on the results of studies that evaluate large groups of patients, in which there has been demonstrated a statistical benefit for the group. The success of this approach is directly related to the biological homogeneity of the chosen study group. Therefore, this approach is inherently sub-optimal for many individual patients, particularly in very heterogeneous disease entities such as diffuse large B-cell lymphoma (DLBL). However, with the recent introduction of high-throughput sequencing, it has become possible to extensively evaluate the biology of individual patient samples, which may eventually be used to transition to an era of true individualized management for cancer patients. Challenges to this approach include the complexity of the technology for clinical laboratories, high cost, and the immaturity of the databases analysis technology that are required to evaluate the results. Fortunately, rapid improvements continue to be made in all of these areas. The primary goal of this project was to explore the feasibility of creating "clinical-grade" evaluation methods toward developing personalized therapies in the near future. Clinical samples from patients with DLBL were used to examine the potential of two platforms, Oxford Nanopore and Illumina company products, for the analysis of complete mRNA transcriptomes since they can be representatives of intracellular biology. I found that the Illumina platform technique is feasible for the goal, while the Oxford Nanopore technology is not. Feasibility was further shown by successful use of the Illumina technology and the analysis method developed, to verify prediction of DLBL aggressiveness as previously determined by an alternate method, considered the "gold standard" in published literature. Finally, mRNA transcriptome data generated from pre-therapy diagnostic and post-

therapy recurrence samples of DLBLs was used to demonstrate that it is feasible to use open-source databases and programs to generate a list of therapeutic candidate proteins and pathways in each individual case. Although this feasibility study was carried out with only small number of patients, it shows that the components may finally be available to consider moving forward. However, further work is required to successful transition individualized tumor evaluation approaches into routine clinical practice.

## Keywords:

Sequencing, Illumina, Oxford nanopore, Diffuse large B-cell lymphoma, DLBL, Whole transcriptome sequencing, Personalized medicine, Cancer.

# Acknowledgment

I would like to thank my Supervisor Dr. Hoyun Lee for his encouragement and constant support over the past two years. This thesis could not have been possible without his guidance, patience, and support. I would also like to thank my committee members: Dr. Rebecca McClure (Cosupervisor) and Dr. Kabwe Nkongolo. I would like to extend my special thanks to Dr. Rebecca McClure for sharing her knowledge, providing helpful advices, and for her exceptional support in writing, reviewing and editing this thesis.

My special gratitude goes to Tyler Kirwan, who always provided valuable guidance and dedicated a lot of his time to teach me the techniques in RNA extraction, library preparation, sequencing, and data analysis. I want to specially thank him for his great help in data analysis using *gnuplot 5.0. patchlevel 6* softwears to generate the heat maps.

I would also like to thank my labmates: Dr. Indeewari Lindamulage, Dr. Vandana Srivastava, Dr. James Knockleby for their help in teaching me different lab techniques and sharing their expertise.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

| | |
|---|---|
| DLBL | Diffuse Large B-cell Lymphoma |
| OS | Overall Survival |
| GCB | Germinal Center B-cell |
| ABC | Activated B-cell |
| RNAseq | RNA Sequencing |
| FF | Fresh Frozen |
| FFPE | Formalin Fixed Paraffin Embedded |
| µM | Micro Molar |
| ml | Milliliters |
| s | Second(s) |
| rpm | revolutions per minute |
| min | Minute(s) |
| µl | Microliter(s) |
| °C | Degree(s) in Celsius |
| PCR | Polymerase chain reaction |
| ng | Nano gram(s) |
| mM | Millimolar |
| pM | Picomolar |
| ECM | Extra Cellular Matrix |

# List of Appendices

# Introduction

Modern techniques for evaluating nucleic acids have clearly demonstrated that each human neoplasm is unique with respect to its genomic variation from normal, at the level of DNA sequence, RNA transcript levels, and expressed proteins in the cells. Within clinical medicine, however, neoplasms are "grouped" into diagnostic categories that have clinical relevance with respect to prognosis and selection of appropriate therapy. These groupings are typically based on what normal cell type the neoplastic cells most resemble, using a combination of features such as clinical signs and symptoms, body location, tissue involved, morphologic pattern, expressed proteins etc. More and more, nucleic acid biomarkers are being used to group neoplasms that are similar not only by the normal cell type that they are most closely resemble, but also by alterations in cell systems that give the neoplasms similar functional characteristics, as these are the features that will be most useful for predicting how the neoplasms will behave and respond to therapy. Typically, translational research studies use clinical diagnostic groupings, with the results of these studies (e.g. whether there is response to a certain drug) being evaluated based on whether there is a statistically significant difference between the group of interest and a selected "control group". The success of this approach is directly related to the homogeneity of the study group of neoplasms such that responses of individual neoplasms within these groups may not be identified, if they deviate from the response of the majority of samples in the group. This "grouping" approach has many other limitations and has clearly been more successful for some groups of neoplasms than others. However, this approach has been considered "the best we have" until very recently, when high-throughput sequencing technologies emerged and began making it possible to evaluate the components of multiple

cellular systems in great detail, simultaneously, for a cost that is approaching acceptability for routine clinical practice. It is these technologies that have allowed very rapid advances in our understanding of basic cell biology at the genomic level and also rapid progression of translational research that is now transforming all areas of medicine (particularly oncology) and ushering in the era of individualized, precision medicine.

In the field of oncology, one goal of individualized precision medicine would be to have "clinical grade" methods to rapidly evaluate many cellular processes of neoplastic cells upon initial patient presentation, such that the spectrum of oncogenic alterations that are unique to each patient's tumor could be identified and appropriate targeted therapy can be selected, or even rapidly designed. This approach would likely be most effective if simultaneous evaluation of the functional status of all of the proteins within the cell were performed; however, such technology for protein evaluation is not yet available. Evaluation of the RNA transcriptome may be the next most informative view of the status of intracellular systems, and advances in sequencing of RNA using high-throughput technology have resulted in methods that may finally be suitable for practical clinical use. High-throughput evaluation of DNA variants is least technically challenging and is already being used in limited capacity in oncology clinics, but DNA information is severely limited by a lack of knowledge regarding how DNA variants (individually, or in combination) ultimately affect the functioning of the proteome.

One of the most challenging areas in oncology is selecting appropriate therapies for neoplasms specially those which are resistant to standard therapy or have recurred following standard therapy. Spear et al. reported that the percentage of cancer patients who actually

respond and benefit from the chemotherapy administrated to them is only 25% [1]. This alarming fact highlights the importance of moving cancer treatment towards more targeted personalized approach. Diffuse large B-cell lymphoma (DLBL) is the most common type of non-Hodgkin lymphoma worldwide, representing 30%-40% of all newly diagnosed cases [2]. DLBL is an aggressive B-cell lymphoma, and it is relatively common in adults, with >50% being resistant to standard therapy or recurring following treatment with standard therapy [3]. Current standard therapy for DLBL patients is the regimen designated as R-CHOP (rituximab, cyclophosphamide, doxorubicin, vincristine, prednisone) [4]. None of the components in this therapy are tumor-specific, with CHOP being a standard chemotherapy cocktail directed at any rapidly dividing cell, and rituximab being only a semi-specific therapy. Rituximab is an antibody targeting CD20, which is expressed in all B-cells, with no discrimination between normal and neoplastic B-cells. The current diagnostic category of DLBL is known to be biologically a very heterogeneous group of neoplasms and many biomarkers have been evaluated for their usefulness in clinically relevant subclassification, with minimal success. As such, no neoplasm-specific genomic biomarkers are currently being used as clinical therapeutics and only small numbers have shown utility even for prognostic stratification within this large group. The latter include increased expression of MYC +/- BCL2 and/or BCL6 at the level of protein detection and *IGH/MYC* +/- *IGH/BCL2* and/or *IGH/BCL6* at the level of DNA detection [5]. For a long time DLBL was considered as one disease until the early 2000s, when it was shown that RNA expression profiling had utility for stratifying DLBLs into at least 2 sub-groups with respect to overall survival (OS). Alizadeh et al.[6] used oligonucleotide capture array technology to evaluate selected RNA transcripts and identified one group resembling normal germinal center B-cells (GCB-type) and a second group resembling blood B-cells that had undergone in-vitro activation

(activated B-cell (ABC)-type). The GCB-type group had a statistically better OS (76%) than the ABC-type group (16%). Similarly, Shipp et al.[7] evaluated RNA transcripts in DLBL by capturing a different selected set of RNA transcripts (also used oligonucleotide arrays) and identified one group with excellent OS ("cured" group - 70% 5 year OS) and a second group with poor OS ("Fatal/Refractory" group -12% 5 year OS).

However, because gene expression profiling has proved to be technically challenging and poorly reproducible in the clinical setting, this type of analysis is not currently used for prognostication of DLBL. Instead, a less-than-optimal surrogate assay is used that includes immunoperoxidase staining of tissue sections to evaluate the expression of 3-6 proteins, and several elaborate interpretation algorithms aimed at achieving "best sensitivity and specificity" for predicting the category that would be obtained using gene expression profiling to separate DLBL into the "cell of origin" groups identified by Alizadeh et al.[6] After years of using this prognostication method in the clinical setting, it has become clear that it is not very reliable; e.g. the GCB phenotype group (good prognostic feature) actually contains some of the most aggressive DLBLs with the poorest survivals, re-emphasizing how conclusions made from analysis of groupings of heterogeneous neoplasms do not translate well for individual patients.

It is clear that better "clinical-grade" techniques and approaches are needed for optimal management of patients with DLBL and there is particular interest in determining whether the concept of real-time individualized tumor evaluation at the genomic level is feasible and could provide for a truly individualized diagnostic, prognostic, and therapeutic approach for these patients. This approach may have particular impact for patients with neoplasms that are resistant

to current therapy or recur following initially successful therapy. Challenges to personalized genomic evaluation of neoplasm, including DLBL, have been many:

1. Techniques for extensively multiplexed evaluation of macromolecules (DNA, RNA, protein) have not been available until recently.
2. As techniques for extensively multiplexed evaluation of macromolecules have emerged, they have still not been suitable for clinical use due to one or more of the followings: too complicated to execute protocols in a clinical lab; data obtained are not reliable or sufficiently reproducible; require too much clinical materials; not compatible with the types of clinical materials typically obtained in clinical work (e.g. paraffin-embedded); required too much materials for the amounts obtainable during clinical work-ups; and/or too costly for routine uses.
3. Insufficient knowledge of "systems biology" such that cellular components being identified as having variations in evaluation of neoplasms, are frequently of unknown function. In addition, there has not been easy access to what knowledge there is.
4. Insufficient knowledge and databases containing information regarding currently available or potential directed therapeutics.

In the last few years, however, there has been substantial improvement in all of the challenging aspects mentioned above. Thus, it is now starting to be conceivable that an individualized, genomic approach to evaluate patients with neoplasms, such as DLBL, may be possible. In particular, high-throughput methods for RNA transcriptome analysis (RNAseq) are particularly attractive to explore. Of course, cellular protein analysis would likely be the best representation of cellular functional states; however, it is not yet amenable.

**Based on this reasoning, the goals of this project were set as follows:**

**1)** Test two very recently developed methods for transcriptome analysis using high-throughput sequencing (RNAseq) to see if either of these methods is feasible to use in a clinical setting at reasonable cost; i.e. are they "clinical-grade" with respect to adequacy for use on routine clinical samples (including paraffin-embedded tissue)?

**2)** Evaluate the preferred RNAseq method for its ability to produce data that can correctly sub-classify DLBL samples into two separate categories that have clinical relevance for prognosis as previously demonstrated using capture array technology.

**3)** Provide preliminary "proof of feasibility" for the use of RNAseq as a method for identifying proteins and/or intracellular pathways that could be candidates for individualized therapy in the clinical setting. This will be done using DLBL samples taken from patients at the time of diagnosis and at post-therapy recurrence, to identify RNA transcripts that show a significant change in expression level post-therapy. Currently available software tools and databases will then be used to identify which cellular pathways contain the predicted altered proteins and investigate whether any of the components or pathways would be amenable to targeted therapy using currently available therapeutics.

# Methods

## RNA extraction from clinical samples

Waste tissue from patients with DLBL was obtained from the Health Sciences North clinical laboratory. Both freshly frozen (FF) and formalin-fixed, paraffin-embedded (FFPE) tissues were used. From each patient, RNA was extracted from at least one sample taken at the time of diagnosis and from at least one sample taken at the time of recurrence. All patients had received the same standard chemotherapy for DLBL: R-CHOP. RNA was extracted using the AllPrep® DNA/RNA FFPE kit (www.qiagen.com), all kit buffers and reagents were used as per the manufacturer's instructions as following: For frozen tissue, excised tumor was disrupted using a mortorized mortar and pestle in lysis buffer and then homogenized using a Qiashredder column, prior to RNA extraction. For FFPE samples, areas of tumor were excised and cut into 10-20 µm thick sections using a scalpel. These sections were placed in a 1.5 ml microcentrifuge tube and deparaffinized by adding 1 ml xylene, vortexed vigorously for 10 s, and spun at full speed (17,000 rpm) in Eppendorf 5415 microcentrifuge for 2 min. (Note: unless specified otherwise, all centrifugations ("spins") were performed at highest speed in Eppendorf 5415 microcentrifuge for 1.5 ml tubes, or BIO-RAD low-speed mini centrifuge for .2 ml PCR tubes). The supernatant was removed; 1 ml of 100% ethanol was added to the pellet to remove residual xylene; the sample was vortexed; and finally, the pellet was spun for 2 min. The supernatant was removed and discarded. With the lid open, the tube was incubated for 10 min at room temperature to eliminate residual ethanol. The pellet was resuspended in 150 µl of buffer PKD, and the tube inverted several times to loosen the pellet. 10 µl of proteinase K was added and

mixed by vortexing, followed by incubation for 15 min at 56°C, and then chilled on ice for 3 min. The sample then was spun for 15 min, and the supernatant (containing the RNA) was transferred into a new 1.5 ml tube, followed by incubation for 15 min at 80°C. Finally, sample was collected after a quick spin.

For RNA clean-up, RNeasy MinElute spin column (www.qiagen.com) was used as suggested by the manufacturer (buffers and reagents supplied in the kit used as per the manufacturer's instructions). Briefly, 320 µl of buffer RLT was added to RNA and mixed by vortexing. 720 µl of 100% ethanol was added and mixed by vortexing. 700 µl of the sample was then transferred to a spin column setting on a 2ml collection tube, and spun for 15 s. The eluate was discarded and the same column was used again as above until the entire RNA sample had been passed through it. The column was given a final wash with 350 µl of FRN buffer, and the final eluate discarded. 10 µl of prepared DNase I stock solution (Dissolved lyophilized DNase I (1500 Kunitz units, provided in the kit) in 550 µl of the RNase-free water) was mixed with 70 µl of RDD buffer by gently inverting the tube. The mixture was directly transferred to the column, incubated at room temperature for 15 min, and then 500 µl of FRN buffer was added to the mixture in the column and spun for 15 s. The eluate (containing RNA) was transferred into a fresh spin column placed at the top of a new 2 ml collection tube, spun for 15 s, and the eluate was discarded. 500 µl of buffer RPE was added to the column, and spun again, this step was repeated twice and the eluate was discarded. The column was then placed in a new collection tube and spun for 5 min. Then the column was placed on 1.5 ml collection tube, 30 µl of RNase-free water added, and RNA was eluted by spinning for 1 min. The extracted total RNA concentration was measured using Nanodrop **(**www.thermofisher.com) and stored at -80°C.

**Synthesis of cDNA from the extracted RNA for RNase P experiment, by using Superscript™ IV First-Strand Synthesis System (www.thermofisher.com)**

To get a total of 20 µl reaction volume, 500 ng of total RNA extract was used for reverse transcription. RNA primer mix was prepared in a small 0.2 ml PCR tube as follow: 50 ng/µl of random hexamers, 10 mM dNTP mix, 500 ng of template RNA, and up to 13 µl of DEPC-treated water were mixed and briefly centrifuged, then heated at 65°C for 5 min. The tubes were then incubated on ice for >1 min. While incubating, RT reaction mix was prepared in a new 0.2 ml tube by combining 5× SSIV buffer (provided in the kit), 100 mM DTT, and 40 units/µl ribonuclease inhibitor. The contents were mixed, spun briefly, and then added to the RNA primer mix. The reaction mix was incubated at 23°C for 10 min and then at 50°C for 10 min. The reaction was stopped by incubating at 80°C for 10min. cDNA concentration was measured using Nanodrop **(**www.thermofisher.com).

**Test the quality of extracted RNA by TaqMan RNase P Detection Reagents kit (www.thermofisher.com)**

To ensure the accuracy of sequencing results, cDNA must be accurately quantified before a sequencing library preparation. *RPP25* (a.k.a. *RNaseP*) is a single copy gene that can be used to accurately quantify amplifiable cDNA or DNA in a sample by comparing it to a standard curve prepared from a sample of known DNA concentration.

Seven tubes of standard DNA serial dilutions were prepared in duplicate, with concentrations of 5, 2.5, 1.25, 0.625, 0.3125, 0.15625 and 0.078125 ng/µl. A dilution of 1:100

ratio of sample cDNA was also prepared in nuclease-free water in duplicate. PCR master mix was prepared by combining the following volumes of reagents per reaction: 10 µl of TaqMan Universal PCR Master Mix (2×) or 20 µl of 1× (AmpliTaq Gold® DNA Polymerase, UP (Ultra Pure), Uracil-N glycosylase (UNG), dNTPs with dUTP, ROX™ Passive Reference, and optimized buffer components), 1 µl of 20× RNase P Primer-Probe mix, and 6.5 µl of nuclease-free water. In a 96 well PCR plate, 17.5 µl of the master mix was added to each well. 2.5 µl of each control DNA dilution were added to "standard" wells. 2.5 µl of sample cDNA dilution was added to each sample well. 2.5 µl of Nuclease-free water was added to negative control wells. Each standard, sample cDNA and negative control were in duplicate. The plate was sealed with a sheet of MicroAmp Optical Adhesive film (Thermo Fisher), centrifuged to ensure all liquid was at the bottom of each well and loaded on a real-time, quantitative PCR machine (Applied Biosystems 7900). Sample cDNA concentrations were determined by comparing to the results of the standard curve DNA samples. Nonamplifiable or poorly amplifiable samples were excluded from further experiments.

**RNAseq Method #1 (Nanopore Technologies)**

The first RNAseq method examined was the commercial platform marketed by Oxford Nanopore Technologies (Oxford, UK). This is considered a 3$^{rd}$ generation high-throughput sequencing system and it is a technique that became available in 2015. It has theoretical advantage over previous methods that would make it an attractive choice for a clinical-grade assay method: Full length RNA sequencing can be performed without fragmentation or initial PCR amplification to create a library, allowing for identification of individual mRNA isoforms;

RNA or cDNA are sequenced directly without intermediate amplification by PCR or branching chain, processes that increase technical complexity and potential errors; no large equipment is needed as sequencing is carried out within an array resembling a data-stick and analyzed directly via consumable that is inserted into any computer; rapid processing time (<24 hours *vs* 3-7 days for second generation sequencing systems). The Oxford Nanopore technology is diagramed in (Figure. 1) and the experiment was performed as per manufacturer's instructions.

**RNAseq Method #2 (Illumina)**

In the fall of 2016, the Illumina sequencer company (illumina.com) released a method called TruSeq RNA Access Library Prep. It allows the evaluation of the entire transcriptome utilizing a standard Illumina technology (second generation sequencing technology), using either fresh or FFPE samples. Importantly, it is amenable with small amounts of RNA typically obtained from clinical samples as it requires only 10-100 ng of RNA. It also can sequence up to 4 samples simultaneously on a single flow cell, which makes the technology cost-effective. This was the first method available for either of the standard high-throughput sequencing machines typically available in clinical laboratories that appeared suitable for the goals of this project (the other is Ion Torrent by Thermo Fisher Scientific). The Illumina sequencing workflow is diagramed in (Figure. 2: a, b) and the library preparation method which includes the conversion of RNA to DNA is described below.

**Sequencing Library Preparation using TruSeq RNA Access Library Prep kit (all the reagents and buffers used are supplied in the kit, Illumina USA, except *AMPure XP beads: Beckman Coulter Canada: A63881, AMPure XP, 60 mL)***

20 ng of total RNA, if isolated from freshly frozen tissue and 100 ng of RNA, if isolated from FFPE tissue were used to prepare the libraries. RNA was diluted in nuclease-free water to a volume of 8.5 µl in a 0.2 ml PCR tube, and 8.5 µl of Elute prime fragment high concentration mix added to the RNA. First strand cDNA was synthesized by adding 8 µl of the FIRST STRAND Synthesis Act D + superscript II mix, and thoroughly mixed by vortexing, followed by a quick spin. The sample was then subjected to cDNA synthesis in the thermocycler, as follows: 25°C for 10 min, 42°C for 15 min, 70°C for 15 min, and then held steadily at 4°C. Second strand cDNA was synthesized by adding 5 µl of resuspension buffer and 20 µl of Second Strand Marking Mix followed by vortexing and a quick spin. The sample was then incubated in the thermocycler at 16°C for 1 h, followed by steadily holding at 25°C.

***AMPure XP clean up (Beckman Coulter Canada: A63881, AMPure XP, 60 mL):***

After generating double stranded DNA, 90 µl of Ampure XP beads was added to it and incubated at room temperature for 5 min, briefly spun, and the tube was positioned on a magnetic stand to precipitate the Ampure beads (magnetic beads). Once the solution was cleared, supernatant was discarded. 200 µl of 80% ethanol was added to wash the sample then the tube was moved back and forth between adjacent wells of the magnetic stand to move the bead pellet within the tubes, the ethanol wash step was repeated twice. The sample was incubated at room temperature for up to 5 min on nonmagnetic stand to evaporate the remaining ethanol, after

which 17.5 µl of resuspension buffer was added, followed by thorough mixing and incubation at room temperature for 5 min. The sample was briefly spun, and then positioned again on the magnetic stand. Once solution was cleared, 15 µl of supernatant was transferred to a new 0.2 ml PCR tube. At this point, the sample could be safely stored at -20 °C for up to 7 days.

***Adenylate 3' Ends:***

*2.5* µl of resuspension buffer was added to the 15 µl sample prepared as above. After 12.5 µl of A-Tailing mix was added, the sample mixed well by briefly vortexing it, followed by quick spin and then incubated in the thermocycler at 37°C for 30 min, 70 °C for 5 min, then held steadily at 4 °C.

***Ligate Adapters:***

2.5 µl of ligation mix was added to a 30-µl sample prepared as above. After 2.5 µl of one unique RNA Adapter index (barcode) was added, the sample was mixed thoroughly, briefly spun, and then incubated in the thermocycler at 30°C for 10 min. 5 µl of kit Stop ligation buffer was added immediately, the sample mixed thoroughly, and briefly spun. The sample was then subjected to two Ampure XP clean up procedures as previously described in ***Ampure XP clean up*** section (page12).

***First PCR Amplification:***

5 µl of PCR Primer cocktail was added to a 20-µl sample prepared as above, followed by addition of 25 µl PCR master mix, mixed thoroughly, briefly spun, and PCR amplified using the following conditions: 98°C for 30 s, 15 cycles of 98°C for 10 s, 60°C for 30 s, 72°C for 30 s,

72°C for 5 min, then finally held steadily at 4°C. The PCR product was cleaned up using Ampure XP beads as described above in ***Ampure XP clean up*** section (page12). At this point, sample can be stored safely at -20°C for up to 7 days.

***Validation and pooling of libraries:***

Each DNA library created as described above was quantified using the nanodrop. 200 ng of each library was then combined to run on the same flow cell (4 samples/ flow cell). The volume was brought up to 45 µl with resuspension buffer to complete the pooled library mix.

***First Hybridization:***

50 µl of Capture Target Buffer 3, and 5 µl of Coding Exome Oligos were added to 45 µl library mix and mixed thoroughly. The tube was incubated in the thermocycler as follows: 95°C for 10 min, and then incubation for 1 min at gradually lower temperature of 92°C, 89°C, 86°C, 82°C, 80°C, 78°C, 76°C, 74°C, 72°C, 70°C, 68°C, 66°C, 64°C, 62°C, and 60°C then held at 58°C. Once the temperature reached 58°C, the sample was incubated at that temperature for 1.5 h, immediately followed by *First Capture* (below) to prevent the sample from cooling down (which could result in non-specific binding).

***First Capture:***

The entire 50 µl of the first hybridization PCR reaction described above was mixed and transferred into a 1.5 ml microtube. 250 µl of Streptavidin Magnetic Beads was added into the sample, vortexed thoroughly, and then incubated for 30 min with occasionally vortexing it to keep the beads suspended. The sample was placed on a magnetic stand and incubated until the

liquid was clear, and then the supernatant was removed and discarded. The beads were resuspended in 200 µl of Enrichment Wash Solution and incubated at 50°C for 20 min. The sample was placed on a magnetic stand, the solution allowed to clear, and the supernatant was discarded. A second wash was performed using the Enrichment Wash Solution following the same steps. During the second 50°C incubation, Elution Pre-Mix was prepared by combining 28.5 µl of Enrichment Elution Buffer1 and 1.5 µl of 2N NaOH (HP3). Once the supernatant was removed after the second wash, 23 µl of Elution Pre-Mix was added to the isolated beads, vortexed thoroughly, and incubated for 2 min. The sample then was placed on the magnetic stand until the solution cleared and 21 µl of the supernatant was transferred to a new 0.2 ml PCR tube, and 4 µl of Elute Target Buffer 2 was added. At this point, sample could be stored at -20°C for up to 7 days.

***Second Hybridization:***

20 µl of resuspension buffer was added to the 25 µl sample prepared by the first hybridization and the first capture step procedures. 50 µl of Capture Target Buffer 3, and 5 µl of Coding Exome Oligos were added to 45 µl library mix and mixed thoroughly. The tube was incubated in the thermocycler as follows: 95°C for 10 min, and then incubated for 1 min at gradually lower temperatures of 92°C, 89°C, 86°C, 82°C, 80°C, 78°C, 76°C, 74°C, 72°C, 70°C, 68°C, 66°C, 64°C, 62°C, and 60°C then held at 58°C. Once the temperature reached 58°C, the sample was incubated at that temperature for 1.5 h, immediately followed by *second Capture* (below) to prevent the sample from cooling down (which could result in non-specific binding).

***Second Capture:***

The entire 50 µl of the second hybridization PCR reaction described above was mixed and transferred into a 1.5 ml microtube. 250 µl of Streptavidin Magnetic Beads was added into the sample, vortexed thoroughly, and then incubated for 30 min with occasionally vortexing it to keep the beads suspended. The sample was placed on a magnetic stand and incubated until the liquid was clear, and then the supernatant was removed and discarded. The beads were resuspended in 200 µl of Enrichment Wash Solution and incubated at 50°C for 20 min. The sample was placed on a magnetic stand, the solution allowed to clear, and the supernatant was discarded. A second wash was performed using the Enrichment Wash Solution following the same steps. During the second 50°C incubation, Elution Pre-Mix was prepared by combining 28.5 µl of Enrichment Elution Buffer1 and 1.5 µl of 2N NaOH (HP3). Once the supernatant was removed after the second wash, 23 µl of Elution Pre-Mix was added to the isolated beads, vortexed thoroughly, and incubated for 2 min. The sample was placed on the magnetic stand until the solution cleared and 21 µl of the supernatant was transferred to a new 0.2 ml PCR tube, and 4 µl of Elute Target Buffer 2 was added. At this point, sample could be stored at -20°C for up to 7 days.

***Capture Sample Clean-up with AMPure XP Beads:***

Capture sample clean-up was done with AMpure XP beads as described previously in ***Ampure XP clean up*** section (page12), followed by a second PCR Amplification. 5 µl of PCR Primer Cocktail, 20 µl of Enhanced PCR Mix were added to the sample, and the DNA was amplified as follows: 98°C for 30 s, 10 cycles amplification at 98°C for 10 s, 60°C for 30 s, 72°C

for 30 s, and 72°C for 5 min, followed by incubation at 10°C. AMpure XP Beads clean-up was performed as previously described in ***Ampure XP clean up*** section (page12).

***Validating the Final Library:***

Quantification was performed using KAPA biosystem quantification kit (www.kapabiosystems.com) following the manufacturer's instructions, which was further confirmed by agarose gel electrophoresis.

***Denature and dilute libraries:***

40 µl of 0.5 nM library was mixed with 40 µl of 0.2 N NaOH, mixed well and centrifuged at 280 ×g for 1 min. The sample was then incubated for 5 min at room temperature to denature double stranded DNA into single strands. 40 µl of 200 mM Tris-HCl, pH 7 was added, mixed, and centrifuged at 280 ×g for 1 min. The sample was diluted to 20 pM concentration by adding 881 µl of prechilled HT1, mixed and centrifuged at 280 ×g for 1 min. The sample was then further diluted to loading concentration of 1.8 pM in 1.3 ml volume by adding 117 µl, 1183 µl of denatured library, and prechilled HT1 respectively. The sample was placed on ice until used.  Illumina NextSeq instrument, NextSeq Platform version NCS v1.3 was used for sequencing.

**Conversion of raw RNAseq data into normalized levels of gene expression to generate RNA expression "heat-maps" (Figure 3).**

Illumina sequenced reads were aligned to the reference human genome (hg19) using the RNA-Seq Alignment Application on Basespace (Illumina Inc.) (STAR aligner algorithm). This

program produces aligned raw reads in the form of bam file format and flags individual reads as properly aligned, low quality, or ambiguous (location cannot be determined because sequence is present in multiple regions of the genome). Reads that were not classified as properly aligned were filtered out. The nature of this library preparation method can cause certain sequence to be overrepresented in the sequencing library simply because they are amplified more efficiently by PCR than others, due to a variety of reasons including differences in length, GC content, and other factors. In order to prevent these kind of PCR artifacts from influencing downstream analysis, all reads that were identified by the STAR aligner as PCR duplicates (having identical length, sequence and mapping location as another properly aligning read) were removed from subsequent analysis. The number of unique, properly aligned reads mapping to each gene were calculated and expressed as raw read counts per gene for each sample.

Raw read counts for each gene within a patient sample were filtered by applying a minimum read threshold of 1 read, which removes all genes without at least 1 read from the analysis. A minimum read threshold of 20 was then applied to each set of paired samples to remove any genes that did not have a minimum of 20 reads in either the initial sample or the recurrence sample. Genes with raw read counts below either of these empirically defined thresholds were considered unreliable and removed from analysis for that sample pair. Raw read counts were then normalized to the number of transcripts from the control gene *ABL1*, to allow comparison of relative gene expression across samples. *ABL1* has previously been shown to be an excellent "house-keeping" gene in lymphoid cells, producing a stable level of transcripts in all cells under a variety of different conditions, and is suitable for this type of normalization[8]. Genes within a pair of samples that had absolute fold greater than the threshold of the minimum fold-changes were determined to be differentially expressed. RNA sequencing data has been

shown to be heteroscedastic, meaning that genes whose read counts differ by orders of magnitude are also expected to have different variances. If not accounted for, this can lead to an increase in Type 1 error. Since the minimal amount of tissue available prevented the use of biological replicates in this study, statistical methods to normalize for variances could not be used. Instead, any differential expression between samples that occurred outside of the two orders of magnitude range from the normalized control gene were considered to be unreliable and filtered out. Hierarchical clustering using Euclidean distance as the similarity measure was performed on the remaining differentially-expressed genes (R version 3.0.2 hclust command). Finally, "Heatmaps" were generated using *gnuplot 5.0, patchlevel 6.*

**Analysis of RNAseq data to assign pre-treatment samples to previously identified, RNA expression-based prognostic categories.**

Sequencing data reads for individual samples were normalized to *ABL1* RNA transcripts: (# of reads per specific gene) × 1000/(# of *ABL1* reads in that sample). Since this analysis compares relative amounts of different genes within the same sample, transcript length had to be taken into account: normalized reads / (size of specific gene in kb). The scale was colored to match that of Shipp et al.[7] (Figure 6a) with lower expression in black, and higher expression in red.

**Analysis of RNA expression levels for potential therapeutic target proteins and cellular pathways.**

DAVID bioinformatics database 6.8 (https://david.ncifcrf.gov/), and Panther classification system (http://pantherdb.org/) were used to identify the genes that are common in a

certain pathway and their biological functions respectively. These databases require the entry of a list of genes. The list of genes generated for each sample with ≥ 5-fold change (≥ 5×) was entered and the results were analyzed. The KEGG_PATHWAY grouping of genes was chosen over other options available by tools in the DAVID system. The results show only the significant, or strongly significant, pathways. DAVID tool uses EASE Score (a modified Fisher Exact P-Value) for gene-enrichment analysis. Panther tool determine p-value by the binomial statistic.

# Results

## Evaluation of the Oxford Nanopore-based RNAseq technique

High quality RNA from the CCRF-CEM cell line was used for a pilot study to examine the feasibility of using the Oxford Nanopore technology for my study. Unfortunately, this technology appeared unable to provide data that would achieve the project's objectives, with the following limitations observed:

- Very low throughput compared to 2$^{nd}$ generation sequencing technologies.
- High error rate.
- High cost per read.
- Inadequate software support to help analyzing the data generated.

Thus, even using high quality RNA from a well-known cell line (it would be worse for actual patient samples, particularly FFPE), it appeared that this system was unlikely going to be adequate for generating "clinical grade" data. Therefore, I decided that it was not worthwhile continuing to test the Nanopore technology for the purposes of this project. It was approximately at this time that the RNAseq method #2 for the Illumina sequencing platform became available. Therefore, I examined the feasibility of this platform, as described in the Materials and Methods section.

**RNAseq analysis using the TruSeq RNA Access library preparation technique and the Illumina NextSeq sequencer.**

As a pilot study, the sequencing method was evaluated by testing two samples of DLBL for which both FF and FFPE materials were available. I examined each sample from both storage methods separately. This was important because most clinical samples are FFPE, and this storage method does not include a nuclease inhibitor which cause the DNA and RNA to get degraded over a long time of storage. Table 1 shows data regarding reads and genes expressed for the two sample pairs following data clean-up and normalization.

Table 1: Read counts and genes expressed in 2 samples taken from same patients (at diagnosis and after therapy relapse), each sample preserved in two different ways, fresh frozen and after formalin fixation with paraffin-embedding (FFPE).

| Sample 1 (at diagnosis) | Fresh frozen | FFPE |
|---|---|---|
| Number of raw reads | 1,943,628 | 3,388,558 |
| Number of expressed genes detected | 14,980 | 14,655 |
| Number of concordant genes | 13,698 | |
| Number of unique genes | 1,282 | 957 |
| **Sample 2 (at recurrence)** | **Fresh frozen** | **FFPE** |
| Number of raw reads | 7,451,942 | 4,430,964 |
| Number of expressed genes detected | 15,547 | 14,800 |
| Number of concordant genes | 14,179 | |
| Number of unique genes | 1,368 | 621 |

Very similar numbers of expressed genes were detected in the samples and the number of concordant genes were also very close to the numbers of detected genes. The presence of unique genes is expected and can be due to the presence of different tumor clones or different amounts of normal cells in different samples. These results confirmed that the RNAseq process had been technically successful. The normalized gene expression data for the entire transcriptome in the samples are shown in (Figure 4). The reproducibility of the expressed genes and concordant genes on the same samples (see also heat map patterns in figure 4) appeared excellent based on the expected intra-sample variation for the RNAseq technique in general. This conclusion was further supported by a comparison of the gene expression profile of each tumor evaluated, which demonstrated a tumor-specific profile that was recognizable for each sample from the others (Figure 5).

I then evaluated the technique's ability to align each sample with one of the 2 prognostic categories identified by Shipp et al.[7]: "cured" group or "fatal/refractory" group (Figure 6a). Since it was known that all neoplasms tested were clinically aggressive, in that they were refractory to therapy or recurred following R-CHOP therapy, it was expected that this analysis, if successful, should clearly align each neoplasm into the "fatal/refractory" group. The results of this analysis are shown in (Figure 6b). Despite the restricted RNA transcript set available for the analysis (see Methods), each pre-therapy DLBL sample did align with the transcript levels and patterns as expected for the "fatal/refractory" group. The clinical features corresponding to each pre-therapy DLBL sample are shown in Table 2.

Currently used clinical prognostic markers for DLBL are reflected in Table 2, as available. These include an immunoperoxidase stain panel that evaluates expressed proteins to determine prognosis per the Hans' algorithm, in which DLBL of germinal center B-cell type (GCB) are considered to have better prognosis than DLBL of non-germinal center type (NGC). Additional prognostic information is obtained by evaluating protein expression of both MYC and BCL2 by immunperoxidase staining, with double-expressers considered to have a very poor prognosis. Finally, evaluation of *MYC*, *BCL2* and *BCL6* at the DNA level is typically done using fluorescence in-situ hybridization (FISH). Disruption of *MYC* (typically indicating a *IG/MYC* translocation) is a poor prognostic feature and disrupted *MYC* with either disrupted *BCL2* or *BCL6* are considered to have an even worse prognosis (aka "double hit lymphomas") [5]. Table 2 shows that some DLBL considered in the "good prognosis" GCB category by Hans' algorithm or that showed no indication of *MYC* disruption, were actually clinically aggressive, re-emphasizing the fact that studies of groupings of heterogeneous neoplasms do not translate well for individual patients.

Table 2: Clinicopathologic features associated with each pre-therapy sample.

| | Sampling at diagnosis | | time to post-therapy sample |
|---|---|---|---|
| Sample 1 (1387) | DLBL (de-novo) | GCB by Hans' algorithm<br>No MYC disruption<br>No MYC IHC done | 43 months |
| Sample 2 (6100) | DLBL (de-novo) | NGC by Hans' algorithm<br>No MYC disruption<br>MYC IHC positive | 26 months |
| Sample 3 (1009) | DLBL (transformed from low-grade) | GCB by Hans' algorithm<br>No MYC evaluation | 25 months |
| Sample 4 (2093) | DLBL (de-novo) | GCB by Hans' algorithm<br>No MYC evaluation | 11 months |
| Sample 5 (706) | DLBL (de-novo) | GCB by Hans' algorithm<br>No MYC disruption<br>MYC IHC not done | 10 months |
| Sample 6 (13084) | DLBL (de-novo) | NGC by Hans' algorithm<br>No MYC disruption evaluation<br>MYC IHC positive | 6 months |
| Sample 7 (948) | DLBL (transformed from low-grade) | NGC by Hans' algorithm<br>No MYC disruption<br>MYC IHC positive | 26 months |

## Further Analysis

Next, I examined pre- and post-therapy samples, FFPE (8 samples) and FF (2 samples) tumor sample pairs from 10 patients using the method described above. Unfortunately, RNA recovery on at least one of the samples in 3 of the pairs of FFPE was too poor for sequencing, leaving complete results on only 7 sample pairs. Transcriptome profile heat-maps for these 7 sample pairs are shown in (Figure 7a-g).

To narrow down the pool of potential targetable proteins, the data was analyzed to identify RNA transcripts, for which expression levels had changed by ≥ 5 times (called ≥ 5× group) between the samples at diagnosis and post-therapy relapse stages (Figures 8a-g). Genes represented in the ≥ 5× group were further evaluated for potential biological functions in the context of cellular pathways using both Panther Classification System tool (www.pantherb.org) and David Functional Annotation Tool (https://david.ncifcrf.gov). A summary of the findings is presented in simplified form in Table 3. Representative screenshots of the types of results obtained from the DLBL cases are shown in Figures 9a-d. Two of the samples (6 and 7) did not have a sufficient number of genes identified in the ≥ 5× group to process through the two analysis tools and therefore were not subjected to these algorithm-based analyses.

Table 3: Some commonly altered gene in different samples. Green boxes show transcripts that were higher in the post-therapy samples. Red boxes show genes that were lower in the post-therapy samples.

| Gene | sample 1 | sample 2 | sample 3 | sample 4 | sample 5 | sample 6 | sample 7 |
|---|---|---|---|---|---|---|---|
| ATRNL1 | | | | | H in REC | | H in REC |
| BGN | H in REC | | | L in REC | | | |
| C4A | H in REC | | L in REC | | | | |
| CCL19 | H in REC | | | | L in REC | | |
| CD6 | | | L in REC | | L in REC | | |
| CDT1 | H in REC | | H in REC | | | | |
| CHIT1 | L in REC | | | | L in REC | | L in REC |
| COL12A1 | | H in REC | | L in REC | H in REC | | |
| COL15A1 | | | | L in REC | | L in REC | |
| COL6A1 | H in REC | | | L in REC | | | |
| COLEC12 | H in REC | | | | | H in REC | |
| CTB-113P19.1 | H in REC | | | L in REC | | | |
| F5 | H in REC | | L in REC | | | | |
| FDCSP | | | L in REC | | | | L in REC |
| HTR3A | H in REC | | | | L in REC | H in REC | |
| JCHAIN | | | | | H in REC | | L in REC |
| LOC100507388 | H in REC | | L in REC | | | | |
| POSTN | H in REC | | | L in REC | H in REC | | |
| PVRL1 | H in REC | | H in REC | | | | |
| SULF1 | | | | L in REC | H in REC | | |
| THBS1 | | | | L in REC | H in REC | | |
| TNC | H in REC | | | L in REC | | | |
| CSMD1 | | L in REC | L in REC | | | | |
| GPR174 | | L in REC | | | | H in REC | |
| ITM2A | H in REC | L in REC | | | | | |
| LINC00707 | H in REC | L in REC | | | | | |
| LINC01609 | H in REC | L in REC | | | | | |
| LOC101927502 | H in REC | L in REC | | | | | |
| LRP1B | | L in REC | L in REC | | | | |
| NKAIN2 | H in REC | L in REC | | | | | |
| NRXN3 | | L in REC | L in REC | | | | |
| OPCML | | L in REC | L in REC | | | | |
| RBFOX1 | | L in REC | L in REC | | | | |
| SORCS3 | H in REC | L in REC | | | | | |

Several observations were made following this exercise. First, the data confirmed that many gene transcripts showed altered expression levels following R-CHOP treatment and relapse, and that the pattern of changes in the expression levels was different in each case. Second, many of the altered transcripts coded for cell components and pathways that are already known to be important for oncogenesis in DLBL and other types of neoplasms. Third, some pathways found to be altered following treatment in this study have not yet been described as important markers in lymphoma. This new finding may provide important new insights into personalize treatment of DLBL.

To extend the proof-of-feasibility exercise, Case 1 was subjected to additional analysis by selecting one high-yield pathway, which I performed literature searches to determine whether specific components of this pathway could be potential targets for therapies with drugs that are already available, or in development. For this exercise, the cytokine pathway was chosen, as it is known to have significance for the pathogenesis of lymphoma, but has not been discussed seriously in most literatures in the context of targeted lymphoma therapies. This exercise required a review of cytokine pathways in general, with specific attention paid to the roles of proteins encoded by genes that had been shown to have altered expressions in the RNASeq analysis of the post-therapy sample in Case 1.

Cytokines are soluble, extracellular small proteins or glycoproteins. They can be grouped into different families, including; chemokines, interferons, interleukins, lymphokines, and tumor necrosis factors. Cytokines are secreted by a broad range of cells in the body. Any given cytokine can be secreted from more than one type of cell, and any one function may be performed by more than one cytokine. They have important roles in fighting infections and other

pathologies, usually by regulating cells' responses for innate and adaptive inflammatory immune responses. Cytokines are also involved in cell differentiation, growth, repair, death, and tissue angiogenesis. They are usually activated by a stimulus, and exert their action by binding to specific receptors on the surface of their target cells.

Normal B-cell development and maturation relies on well-regulated interactions with other immune cells and stromal cells. Interactions between lymphoma cells with their microenvironment are also essential for cancer development and metastasis.[9] Tumor microenvironment components are not only part of the body's antitumor inflammatory response, but they also play a critical role in enabling cancer progression[10]. Cytokines and their receptors facilitate the crosstalk between normal and neoplastic cells, and the presence of cytokines in the tumor microenvironment contributes to cancer pathogenesis[11]. In addition, compelling epidemiological data shows that unsolved body immune reactions, specifically diverse forms of chronic inflammation, promote malignant transformation and cancer development[11]. Cancer cells use host derived cytokines that normally function to promote growth, constrict apoptosis and assist with invasion and metastasis to their advantage.

Since lymphoma cells are immune cells, they co-opt the trafficking and homing of normal immune cells to locate a supportive environment for their growth. Lymphoma cells express cytokine receptors, and the pattern of expression of these receptors correlate with the site of their metastasis. Chemokine receptors CCR7, CXCR4 and CXCR5 are involved in homing and trafficking within lymph nodes, and it has been observed that they are highly expressed on lymphoma subtypes with widespread nodal metastasis. These receptors are less expressed on lymphoma subtypes that remain localized.[10] Cytokines-cytokine receptor pathway components

represent an attractive promising target in oncology therapeutics, and are currently a hot topic in clinical trials.

The following discussion focuses on cellular functions of mRNAs that showed increased expression following therapy in Case 1, of which components were identified by DAVID tool. Current data suggests that any or all of these cytokines/cytokine receptors could be potential targets for directed personalized therapies.

**C-C Motif Chemokine Ligand 19 (CCL19): showed (8.5)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than in the initial sample:**

CCL19 is a homeostatic chemokine, it is expressed in the secondary lymphoid organs. It regulates lymphoid cells, and has a role in Dendritic cells (DCs) homing. It could promote inflammation, and it has been associated in various inflammatory diseases and infectious disorders. Its receptor is CCR7.[12]

CC cytokines fall under the chemokine superfamily, which is further classified into subgroups, CC, CXC, CX3C and C. Members of this super family promote cytoskeleton rearrangement, adhesion to endothelial cells, and directional migration, with involvement in many different stages of tumor development including initiation, growth, and progression.[13] It has been previously reported that this system can be hijacked by epithelial cancer cells and may facilitate the dissemination of tumor cells. Chemokines and their receptors have also been identified as possible contributors to the metastatic process[14], and many studies have shown that CCL19 increases cell proliferation in many different cancers.[13]

DCs play an essential role in the initiation and regulation of antigen specific immune responses, and they have the ability to initiate antitumor immune responses. Upon encountering antigens, immature DCs take these antigens and process them. They then undergo a maturation process with expression of CCR7, which directs these cells toward their ligands CCL19 or CCL21. These ligands are usually present in T-cell enriched areas in the secondary lymphoid organs. Once DCs present the processed antigens to T-cells, they start to differentiate and attack their targets.[15] In many different cancers, the DCs have impaired maturation, leading to decreased stimulatory effects on T-cells. Hwang et al.[15] conducted a study using a single cell-based analysis in a 3D microfluidic device on several breast cancer cell lines. They observed that breast cancer cells release soluble factors, increasing CCL19-induced directional persistence of DCs. The triple negative breast cancer cells facilitated this movement by upregulating the JNK/c-Jun signaling pathway. They also noticed that triple negative breast cancer cells upregulated DCs secretion of pro-inflammatory cytokines. When T-cells are induced by DCs, they become extremely proliferative and resistant to activation of induced cell death, and their secretion of pro-inflammatory cytokines increased. The high levels of pro-inflammatory cytokines may result in creating a host inflammatory environment which may promote tumor growth.

**C-X-C Motif Chemokine Ligand 12 (CXCL12): showed (12.11)-fold higher expression in the recurrence sample (In relative to *ABL1* gene) than the initial sample:**

CXCL12 (also called SDF-1) binds to its receptor CXCR4, the downstream effects of this activation ultimately affect many cellular functions including inflammation response, tumor progression, metastasis, vasculogenesis, and the mobilization of hematopoietic stem cells. [16]

CXCL12 is crucial for normal B-cell growth, and it is expressed in normal tissues and serum. It is also expressed by stem cells, endothelial cells, multiple immune cells, stromal fibroblasts, and cancer cells. Levels of CXCL12/CXCR4 were reported high in patients with many solid tumor types such as breast, gastric, pancreatic, ovarian, cervical, and carcinoma of the oral cavity. Furthermore, CXCL12/CXCR4 levels were higher in patients with advanced stages of the B-cell neoplasm chronic lymphoblastic leukemia patients (CLL) than in patients with lower stages. CXCL12/ CXCR4 can affect cancer by two mechanisms: CXCL12 can exert a direct autocrine effect enhancing cancer cell progression and angiogenesis; and it can also attract cancer cells expressing CXCR4 to CXC12-expressing organs to initiate metastasis in those organs. Guo et al.[17] has suggested that CXCL12 can activate the NFκB pathway, which can suppress apoptosis. CXCR4-positive inflammatory, vascular and stromal cells can be attracted to the tumor mass if the tumor had high levels of CXCL12. These cells then release growth factors, cytokines, chemokines and pro-angiogenic factors providing the perfect microenvironment for cancer development.

In Case 1 recurrence sample, high expression of CXCL12 and other cytokines and their receptors were observed. Components of the NFκB pathway were also identified by DAVIDS tool to be significantly upregulated in this case. These observations are consistent with those of Guo et al.[17]

**C-X-C Motif Chemokine Ligand 13 (CXCL13): it had (9.1)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than the initial sample:**

CXCL13 is a B-cell chemoattractant and it is highly expressed in secondary lymphoid organs by T-follicular helper cells, DCs, and stromal cells. It promotes the migration and mobility of B-cells, which express its receptor CXCR5.[18]

In a study of breast cancer patients, high expression of CXCL13 was associated with poor prognosis[19]. Singh et al.[20] reported significantly elevated levels of CXCL13 in the serum of prostate cancer patients compared to normal healthy donors, benign prostatic hyperplasia, and high-grade prostatic intraepithelial neoplasia. They also reported that CXCL13 was a better predictor of prostate cancer than prostate-specific antigen. CXCL13 and CXCR5 high expression was also correlated with the development, metastasis, and recurrence of colon cancer[21].

Rubenstein et al.[22] conducted a study to compare the levels of CXCL13 and IL-10 in central nervous system fluid from lymphoma patients and those with inflammatory and degenerative neurologic diseases. Increased levels of CXCL13 and IL-10 were specific for primary and secondary central nervous system lymphoma, and associated with poor outcome. Thus, the authors suggested the use of CXCL13 and IL-10 as a diagnostic marker in primary CNS lymphoma.

**Tumor Necrosis Factor Ligand Superfamily Member 7 (CD70): it had (15.6)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than the initial sample:**

CD70 is a cytokine that falls under the tumor necrosis factor (TNF) ligand family and binds to its receptor CD27. This binding plays an important role in immune regulation. It is

solely expressed and regulated upon immune activation by an antigen. It is expressed on activated DCs, B, T-cells, and natural killer (NK) cells. CD70 is required by epithelial and other cell types during malignant transformation. [23]

Normally, CD70 activates the NFκB and c-Jun kinase pathways upon binding to its receptor CD27. The cytoplasmic residues of CD27 then bind to TNF receptor-associated factors such as TRAF2 and TRAF5, ultimately leading to proliferation, differentiation, and survival. CD70 expression has been reported in both hematological and solid tumors; in particular, in lymphomas, renal cell carcinoma, nasopharyngeal carcinoma, and Epstein–Barr virus-induced carcinomas. High expression of CD70 in B-cell lymphoma, renal cell carcinoma, and breast cancer is associated with poor prognosis. CD70 is expressed in the primary tumors, and up to 100% stable expression of CD70 is found in metastatic patient-derived tissues. Aberrant epigenetic of the CD70 promotor region primarily demethylation has been associated with constitutive expression of CD70 in large B cell lymphoma. [24]

Recently, CD70 and CD27 have emerged as valuable targets for immunotherapy.

**TNF Receptor Superfamily Member 21(TNFRSF21): it had (19.4)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than the initial sample:**

TNFRSF21 also called death receptor 6 (DR6) is a member of the death receptor family, which belongs to the tumor necrosis factor receptors super family. This receptor has the ability to induce apoptosis by activating the NFκB pathway and the mitogen-activated protein kinase 8. DR6 is expressed in many tissues but with higher expression in lymphoid organs, heart, brain and pancreas. it plays a role in regulating immune responses. [25]

The mechanism of action of TNFRSF21 is still not fully understood and its ligand is still unknown. However, it is known that TNFRSF21 consists of two domains, an extracellular cysteine rich domain and an intracellular death domain that induces apoptosis. High levels of TNFRSF21 have been observed in the late stage ovarian cancer, prostate cancer, breast cancer and a variety of solid tumor cell lines. [26],[27],[25]

Yang et al.[26] demonstrated that TNFRSF21 can have a role in either cancer cell survival or death, depending on the microenvironmental conditions. They also suggested that TNFRSF21 has a potential role in the tumor microenvironment enhancing angiogenesis and facilitating tumor growth. Upon knocking down TNFRSF21 in mouse melanoma, tumor growth was inhibited by suppressing the expression of VEGF-A, PDGF-β, VEGF-D and PDGFR-α (blood vessel formation related factors).

**Interleukin-10 (IL-10): it had (8.4)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than the initial sample:**

IL-10 is an anti-inflammatory cytokine, it plays an important role in the regulation of immune responses, and preventing autoimmune diseases. It is widely expressed by both innate and adaptive immune system cells.[28]

IL-10 expression may be regulated pre-and/or post-transcription. The role of IL-10 in cancer development and progression is controversial but multiple studies have reported a direct correlation between IL-10 expression in serum or tumor and poor patient prognosis. Cancer cells may express IL-10 to escape the immune surveillance. If the cancerous cells expressed IL-10R, the production of IL-10 in the tumor microenvirmoent will mainly act as cancer development

promotor. The ability of IL-10 to downregulate MHC II in tumor cells provides an immunosuppressive environment facilitating tumor escape from the immune system. Other studies found that IL-10 had antitumor activity by stimulating NK-cells and cytotoxic T-cells to kill cancer cells. The presence of other cytokines in the microenvironment can affect IL-10 function and its positive or negative effects on cancer progression.[29]

**IL-6 Receptor Subunit Alpha (IL6R): it had (10.5)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than the initial sample:**

It belongs to class I cytokine receptor family. It is a multi-chain receptor complex. It has ligand binding components and signal transduction components. The ligand binding part can be found in both soluble or membraned bounded forms, while the signal transduction part is a glycoprotein.[30]

When IL-6 binds to IL-6R family members, any or all of the JAK–STAT, MAPK, PI3K pathways are activated, ultimately causing changes in the transcription of genes involved in proliferation (c-MYC, cyclin D1), angiogenesis (VEGF, notch3), metastasis (MMP9, CXCR4, CXCL12), chemoresistance (MDR1, GSTpi) and survival (Bcl-2, Mcl-1).[31]

**Interleukin 7 Receptor (IL7R): it had (8.1)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than the initial sample:**

IL-7 binds to its receptor IL-7R generating essential signals for normal T-cell development and homeostasis. Zenatti et al. [32] reported that IL7R mutations play an important role in tumor formation and progression. It has specially an active role in T-cell leukemogenesis.

Pro-tumor effects of IL-7 and its receptor are thought to occur by preventing apoptosis. A study suggested that IL-7 and its receptor upregulates cyclin D1 and Bcl-2, decreases P53 and BCL2 associated X protein. IL-7 may increase proliferation and lymphovascular formation. It has been reported as an oncogene in T-cell acute lymphoblastic leukemia.[33]

**Lymphotoxin Alpha (LTα): it had (8)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than the initial sample:**

LTα previously known as TNFβ, belongs to the TNF family. It is a soluble homotrimer that can attach to the cell surface only when forms heterotrimers with lymphotoxin beta. It is structurally similar to TNFα. It exerts its biological function by binding to its receptors TNFR1 and TNFR2. It is expressed by CD4+ T helper type 1, CD8+ T cells, NK cells, B cells, and macrophages. It has a role in lymphoid organ development, immune system function, inflammation, host defense, and maintenance of lymphoid microenvironment. Previous studies reporting the functions of this cytokine used mice lack LTα gene, and the applicability of LTα roles in human is less clear. [34]

**TNF Superfamily Member 12 (TNFSF12): it had (20.9)-fold higher expression in the recurrence sample (in relative to *ABL1* gene) than the initial sample:**

TNFSF12 is also known as TWEAK and it is a ligand for FN14/TWEAKR. It has a role in numerous cellular activities involving differentiation, migration, proliferation, inflammation, angiogenesis, and apoptosis. It is expressed as a transmembrane protein, then it is proteolytically cleaved into the soluble active cytokine. TWEAKR can be found in various tissues, it is highly expressed in the context of injury and tissue regeneration.[35]

Studies have shown that both TNFSF12 and its receptor have high expression in inflammatory cells and cancer cells. The cellular activities promoted by TNFSF12, including cell survival, migration, angiogenesis, proliferation, and differentiation inhibition, are linked to tumorigenesis.[36]

In summary, published data indicate that cytokine pathways in general may be targets for manipulation in many different forms of cancers, including B-cell lymphomas. There is also evidence in literatures that the specific components of the pathways identified as up-regulated following therapy in Case 1 would be worth investigation as possible targets of directed therapy in this specific DLBL. If the methods used in this study were to be clinically useful, however, directed therapies for these specific components discussed above or other components of the pathways identified as altered in this sample, would need to be clinically available. To gain some insights into this aspect, some of currently available drugs targeting the cytokine pathway are shown in Table 4. Since the levels of mRNA of these potential target genes are upregulated by five-folds or more in Case 1, these drugs, singly or in combination, can be effective against DLBL.

Table 4: Potential targeted therapies available for components of the cytokine pathway(s). mAb denotes monoclonal antibody.

| Gene | Drug | Phase | Mechanism | Reference |
|---|---|---|---|---|
| **CXCL13** | MAb5261 | Preclinical | • mAb*<br>• Blocks CXCL13 | Klimatcheva E et al. BMC Immunology 2015;16:6.[37] |
| **CD70** | SGN-CD70A | Phase I trials | • mAb-toxin conjugate<br>• Binds CD70 & internalized toxin inhibits DNA replication | Jacobs J et al. Pharmacology and Therapeutics 2015;155:1-10.[38] |
| | AMG 172 | Ph<br><br>ase I trials | • mAb-toxin conjugate<br>• binds CD70 & internalized toxin disrupt microtubules | Jacobs J et al. Pharmacology and Therapeutics 2015;155:1-10.[38] |
| | ARGX-110 | Phase I trials | • mAb<br>• Neutralizes CD70<br>• Allows Ab-dependent complement toxicity and phagocytosis | Jacobs J et al. Pharmacology and Therapeutics 2015;155:1-10.[38] |
| **IL10** | B-N10 | Preclinical | • mAb<br>• Neutralizes IL-10 | Llorente L et al. Arthritis and Rheumatism 2000;43:1790-1800[39] |
| **IL-6R** | Tocilizumab | Approved | • mAb<br>• Inhibits binding of IL-6 to IL-6R | Hunter C and Jones S. Nature Immunology 2015;16:448-457.[40] |
| | Sarilumab | Approved | • mAb<br>• Inhibits IL-6R | Hunter C and Jones S. Nature Immunology 2015;16:448-457.[40] |
| | Siltuximab | Approved | • mAb<br>• Inhibits binding of IL-6R | Hunter C and Jones S. Nature Immunology 2015;16:448-457. 2015.[40] |
| **IL-7R** | OSE-127 | Preclinical | • mAb<br>• Inhibits IL-7R | http://ose-immune.com/en/portefeuille-de-produits/effi-7/ |
| **TNFSF12** | RG7212 | Phase I trials | • mAb<br>• Inhibits TNFSF12 | Cheng E et al. Frontiers in Immunology 2013;4:1-13.[41] |
| | PDL192 | Phase I trials | • mAb<br>• targets fn14 (TNFSF12 receptor) | Cheng E et al. Frontiers in Immunology 2013;4:1-13.[41] |
| | Fn14-TRAIL | Preclinical | • Fn14-TRAIL chimeric molecule<br>• Blocks TNFSF12 | Cheng E et al. Frontiers in Immunology 2013;4:1-13.[41] |

# Discussion, Conclusion, and future work

Through the work presented here, all three goals established at the on-set of this project have been achieved as described below:

1. An RNASeq method that can achieve "clinical-grade" transcriptome analysis has been identified.
2. The RNASeq method was used successfully to classify patients with DLBL into correct subgroups of clinical prognostic categories as defined by prior "gold-standard" RNA transcription studies using “research-grade” techniques.
3. A proof of feasibility study has been successfully carried out to prove the concept that this RNASeq method can identify mRNA transcripts that are significantly altered post-therapy relapse in DLBL. Furthermore, publicly available analysis tools can be used to demonstrate how the corresponding proteins may be important for functions in normal and neoplastic cell systems so that promising targets for individualized therapy can be identified. Published data support this conclusion as pharmaceuticals that are capable of altering components of the pathways identified by this study with patient’s samples have also previously identified by other methods such as nucleotide microarrays and gene expression analysis by RT-PCR. Since these pharmaceuticals are already available to use at clinics, their efficacy and specificity associated with biomarkers identified by this study may be examined in a clinical setting.

Clearly, the work presented here is only preliminary. However, this type of proof of feasibility study is a critically important step toward the ultimate goal of truly individualized

cancer therapy. As far as I know, no similar study has thus far been published in the scientific or clinical literatures. Many studies have evaluated groups of DLBL samples either at the diagnostic or after therapy stage, but studies reporting a detailed investigation at individual neoplasms at pre-and post-therapy stage appear absent. The importance of this approach (investigating individual neoplasms at pre-and post-therapy) was reinforced by the encouraging data shown in this study. Although several intracellular pathways may be altered post-therapy in most cases, each individual patient showed also a unique combination of altered pathways or components in the single pathway. In addition, I have also identified, several alterations in pathway components that have not yet been identified as important in DLBL, but are being investigated in other types of neoplasms and were reported to have a diagnostic or prognostic values such as TNFRSF21, TNFSF12, CCL19, and some other components of different pathways.

As discussed in the introduction, previous studies on DLBL transcriptomes have been carried out using oligonucleotide arrays containing pre-selected targets based on prior knowledge of pathways thought to be relevant to lymphoma. Other studies used similar pre-selected panels for some relevant genes. The technology in the current study differs from prior work in at least two ways, which can be advantageous for clinical use:

1. The entire transcriptome (all known coding gene products) are theoretically detectable using the Illumina assay, so the detection of transcripts is not restricted to a subset already described in the literature. This should allow for the detection of transcripts and pathways that may be altered and have not been previously identified in drug-resistant/recurrent DLBL.
2. While prior studies used oligonucleotide arrays to capture targeted mRNAs, the Illumina method uses sequencing of mRNA to detect the transcripts that are present. High-

throughput sequencing is a more powerful technology than oligonucleotide arrays, not only because it is less restrictive (as discussed above) but also because it is more flexible. In addition, it can be done using very small amounts of RNA and it is cost effective for routine use. Above all, the utility of this method has been “verified” by results obtained from clinical and oligo array-based studies.

Although adoption of high-throughput sequencing technology has been rapid and is generating a wealth of basic cell biological data, the development of databases and software to store and access this data in a useful manner has lagged behind. Only recently have these tools been available, and it is necessary to determine if they are sufficiently developed to make the approach piloted in this small study feasible for clinical use. The experience obtained during this study suggests that most products are still quite primitive with respect to the amount of information contained, and quite "clunky" with respect to the ease of use for the purpose required. However, these vast amounts of data/information should eventually be useful for practical purposes, and this aspect moves forward very rapidly. The work described in this thesis is part of this effort, and has clearly demonstrated that the analysis of transcriptome with high throughput RNAseq-based study is feasible to identify potentially disrupted cellular pathways that can eventually lead to the development of therapeutic targets for each individual cancer patient. Thus, it was concluded that this feasibility study project was successful, and that it is worth continuing to pursue this approach using DLBL as a model.

A more in-depth study on this project has been limited due to the short time-frame and insufficient research budget. Assuming these limitations can be addressed, the following work may be considered to furthering this work:

1. Further ensurance of the reproducibility of the RNASeq method would be needed. Although my data generated using relatively small number of samples appear to be reproducible, more samples and more extensive evaluation of the transcript variations may be needed to further ensure the reliability of data.
2. More work is needed to determine if the sensitivity of the transcript detection is adequate for the most clinically relevant transcripts and that their alterations are relevant to tumor development, progression, resistant to therapies, and the use of appropriate therapies. It has been suggested that some of the most important transcripts are found at very low levels; therefore, it may be difficult to detect unless a highly sensitive method is used. The sensitivity of the assay was not systematically examined in this project. Further refinement may be needed to obtain truly reliable data.
3. It would be useful to evaluate and normalize the sequencing data using different algorithms, not only one, as done in this study. There are several published algorithms that maybe considered to use in the future, such as culling of reads and using different read normalization methods as mentioned in Li et al.[42]
4. Evaluation of significantly altered transcripts could be greatly expanded. In this study, only those transcripts that showed at least 5× alterations in mRNA levels, and only a single pathway in one case (Case 1) has been evaluated in detail. Many studies have suggested that alterations as low as 2× could be significant and that transcript sets, albeit large, may be clinically relevant. It may be also possible to evaluate for only those transcripts that are even more substantially altered (e.g. 10×) to identify truly significant alterations that are directly relevant to the clinical setting. In any cases, the manual

process used in this study is very labor intensive and future work should be streamlined by development of an automated algorithm.

5. It would have been interesting to more extensively compare the data obtained with prior transcriptome analyses of DLBL patients with respect to known, biologically and clinically relevant subgroups. Although this was not the primary goal of this project, the ability to stratify DLBL at the time of diagnosis into relevant prognostic subgroups using the current “gold-standard" RNA evaluation approach has been elusive for clinical laboratories. It would be exciting to extend this work into a true validation of this type for clinical assay development.
6. Much more time could be spent evaluating the data using tools designed to identify clinical relevance of the transcripts showing significant alterations following therapy. There are more tools available, each with advantages and disadvantages, and finding the correct combinations of tools and building an algorithm that could be validated for real clinical use would be a large project in itself.
7. The testing of potential targeted therapies using an *in vitro* system would be a critical next step toward proving feasibility of this approach to individualized medicine. Although the time frame of this project did not allow for prospective collection of fresh diagnosis/recurrence sample pairs that could be stored as viable cells for later studies, this type of tissue would be useful. The goal would be to perform the RNASeq transcriptome experiment as done in this project on both samples, and then develop a system through which these cells could be grown *in vitro* (or in animals) long enough to carry out experiments to determine the most effective therapies. Although this can be challenging,

it is certainly doable. If successful, the outcome of cancer therapy can be dramatically improved.

8. The analysis of the cytokines pathway showed that these cytokines and their receptors play an important role in DLBL relapse. chronic prolonged inflammation promotes cancer progression and make it more resistance to therapy. The components of this pathway and their effects were studied in other types of cancer, but they weren't sufficiently studied in DLBL since it's a cancer of immune cells and the inflammation was considered a normal result of the stress of the immune cells. The already available drugs which target some of these cytokines and their receptors were developed in studies considering other types of cancer, it is worthwhile studying them in DLBL in-vivo and in-vetro models.
9. This type of studies has a great impact on the clinical management. If a low-cost, reliable sequencing method was developed and routinely practiced in the clinics, it will increase the accuracy of the diagnoses and prognosis, increases the cure rates, and it will reduce the costs of the diagnostic tests. It will also eliminate the unnecessary use of the expensive and toxic chemo agents that might harm the patients instead of curing them. These studies are important to lay the foundation for the development of personalized medicine and they increase the pace of the movement towards it.

# Figures and legends

Figure 1: MinION: Oxford Nanopore's "3rd generation" Sequencer. (1, 2) Protein nanopores are embedded in the membrane. (3, 4) A current passes through the pore, making a voltage across the membrane. As analyte passes through the pore, there is a characteristic disruption of current that can be measured, and the signature used to identify what nucleotide is passing through the pore (DNA base, or RNA base). The entire array is embedded in a "stick" that can be inserted into any computer following the run for direct analysis.

Image source: John MacNeill, http://www2.technologyreview.com/article/427677/nanopore-sequencing

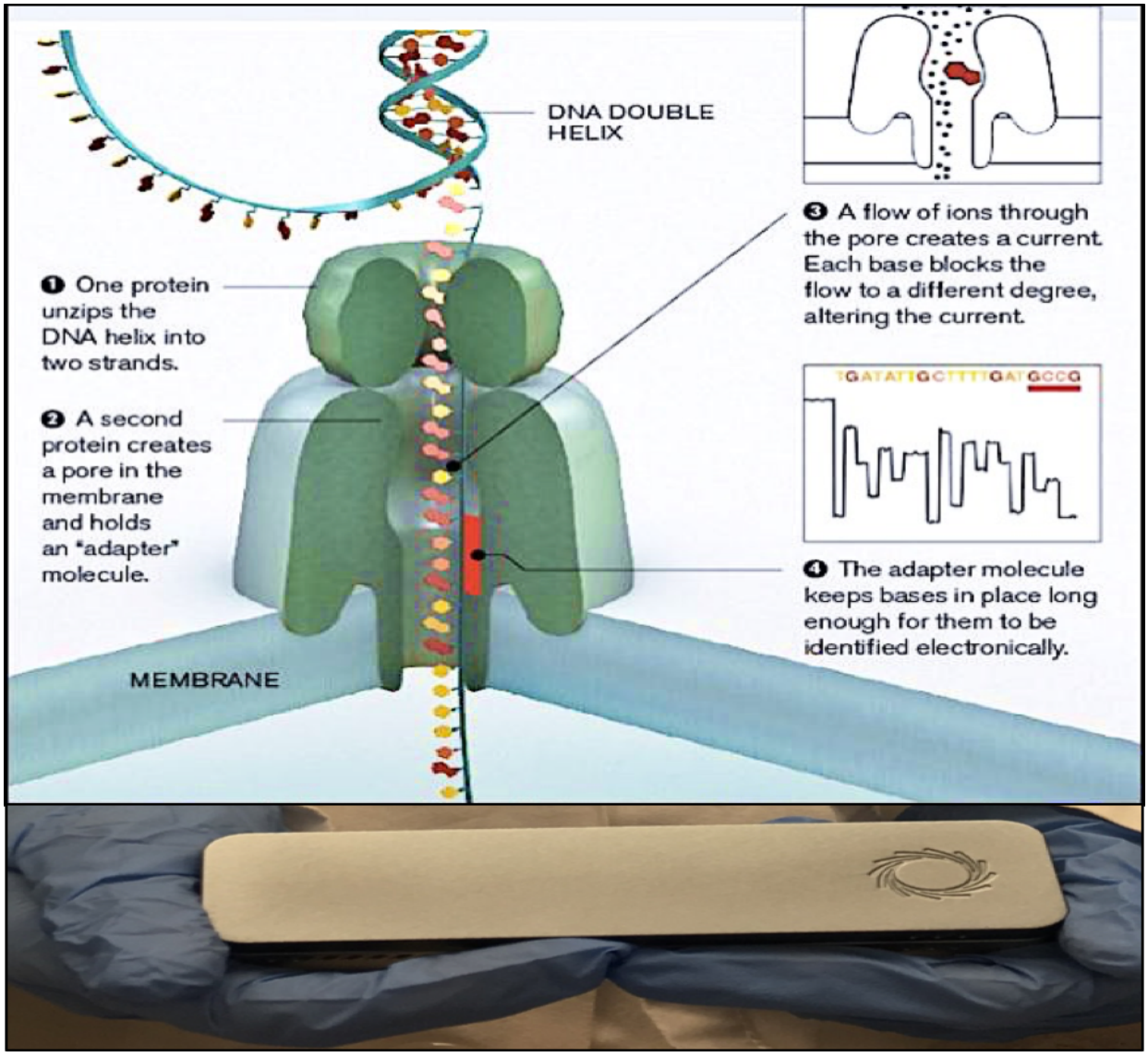

DNA DOUBLE HELIX
1 One protein unzips the DNA helix into two strands.
2 A second protein creates a pore in the membrane and holds an "adapter" molecule.
3 A flow of ions through the pore creates a current. Each base blocks the flow to a different degree, altering the current.
4 The adapter molecule keeps bases in place long enough for them to be identified electronically.
MEMBRANE

Figure 2a: Illumina Sequencing Technology workflow. (1): A library of DNA is created (typically via targeted PCR or fragmentation) and Illumina sequencing "ends" attached to each fragment. (2): Each fragment is secured to flow cell in a unique location via attachment to stationary oligonucleotide (primer) with sequencing homology to one of the sequencing ends on each fragment. (3, 6): A technique called bridge amplification allows for amplification of each fragment into a cluster of identical fragments at each flow cell location, as is required to obtain sufficient signal for detection during the sequencing reaction steps. (https://www.illumina.com/).

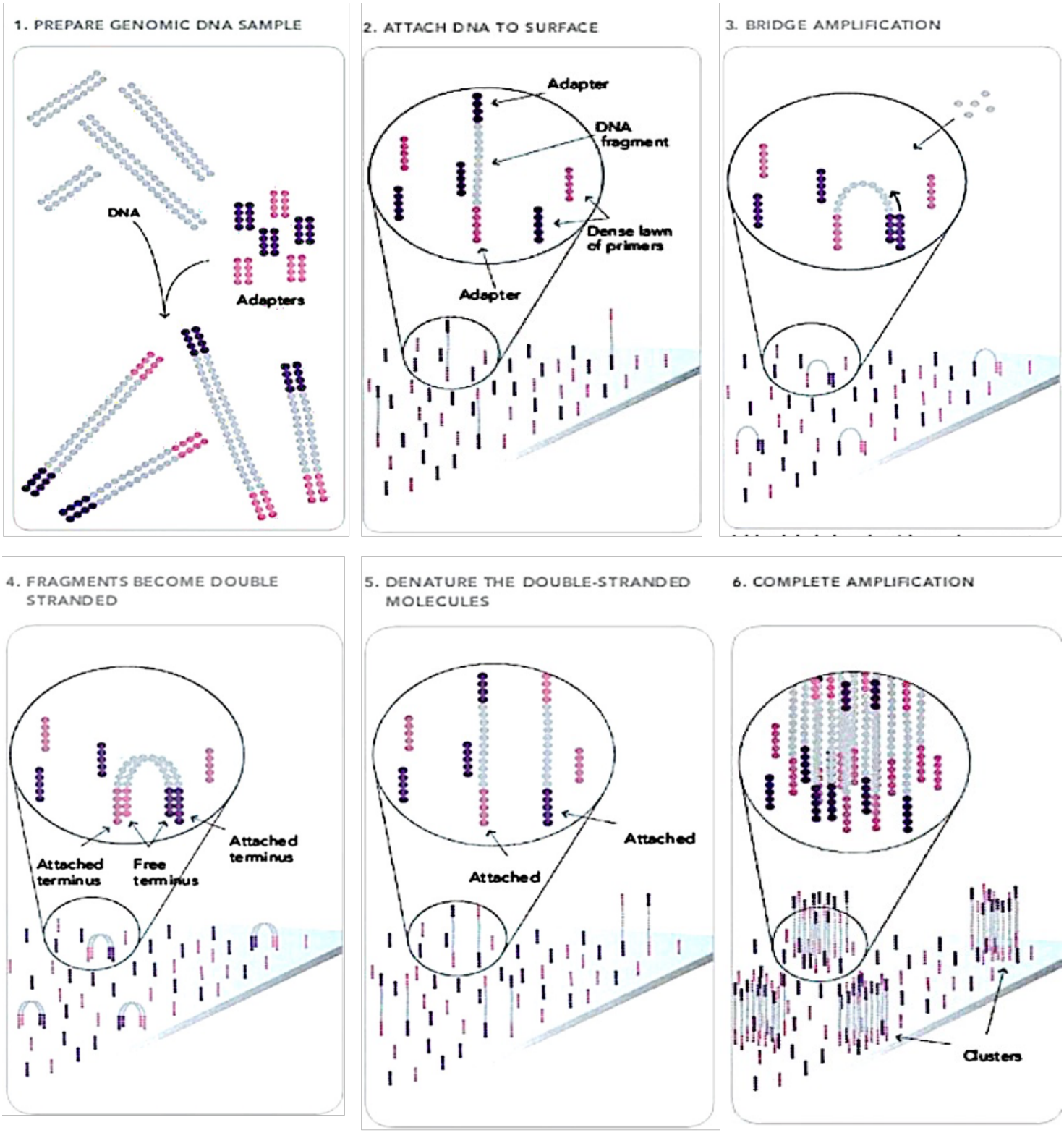
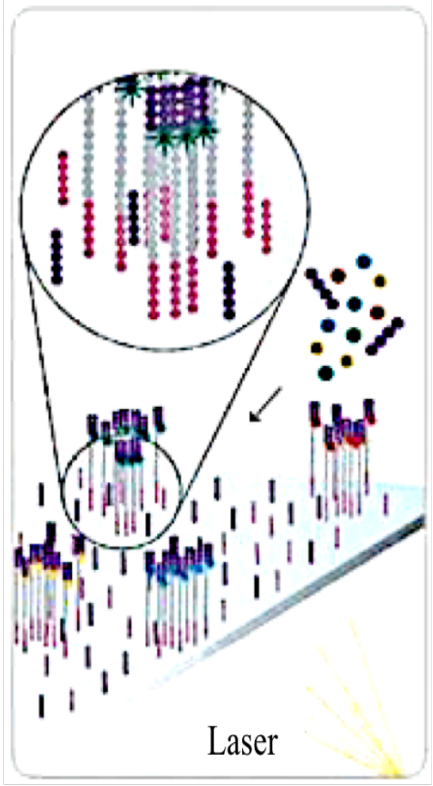

1. PREPARE GENOMIC DNA SAMPLE
DNA
Adapters
2. ATTACH DNA TO SURFACE
Adapter
DNA fragment
Dense lawn of primers
Adapter
3. BRIDGE AMPLIFICATION
4. FRAGMENTS BECOME DOUBLE STRANDED
Attached terminus
Free terminus
Attached terminus
5. DENATURE THE DOUBLE-STRANDED MOLECULES
Attached
Attached
6. COMPLETE AMPLIFICATION
Clusters

Figure 2b: Illumina sequencing technology workflow. Sequencing is performed by cycling of fluorescently labeled basis (one color per base, one base per flow) onto the flow cell. The sequencing occurs from the top of each fragment down, with base binding, if complementary on each cycle. In each cycle, a camera under the flow cell detects whether binding has occurred in each location and if so, knows which base bound, based on which nucleotide is being flowed in the cycle. After each cycle, the nucleotide is washed off and the next nucleotide flooded onto the array. (https://www.illumina.com/).

7. Determine first base




8. Image first base


9. Determine second base

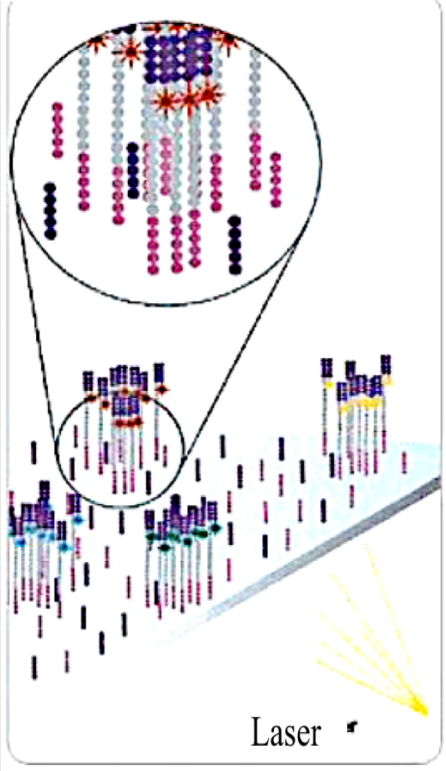


Figure 3: Sequencing read preparation and normalization flow-chart, explained in detail in the Materials and Methods section.

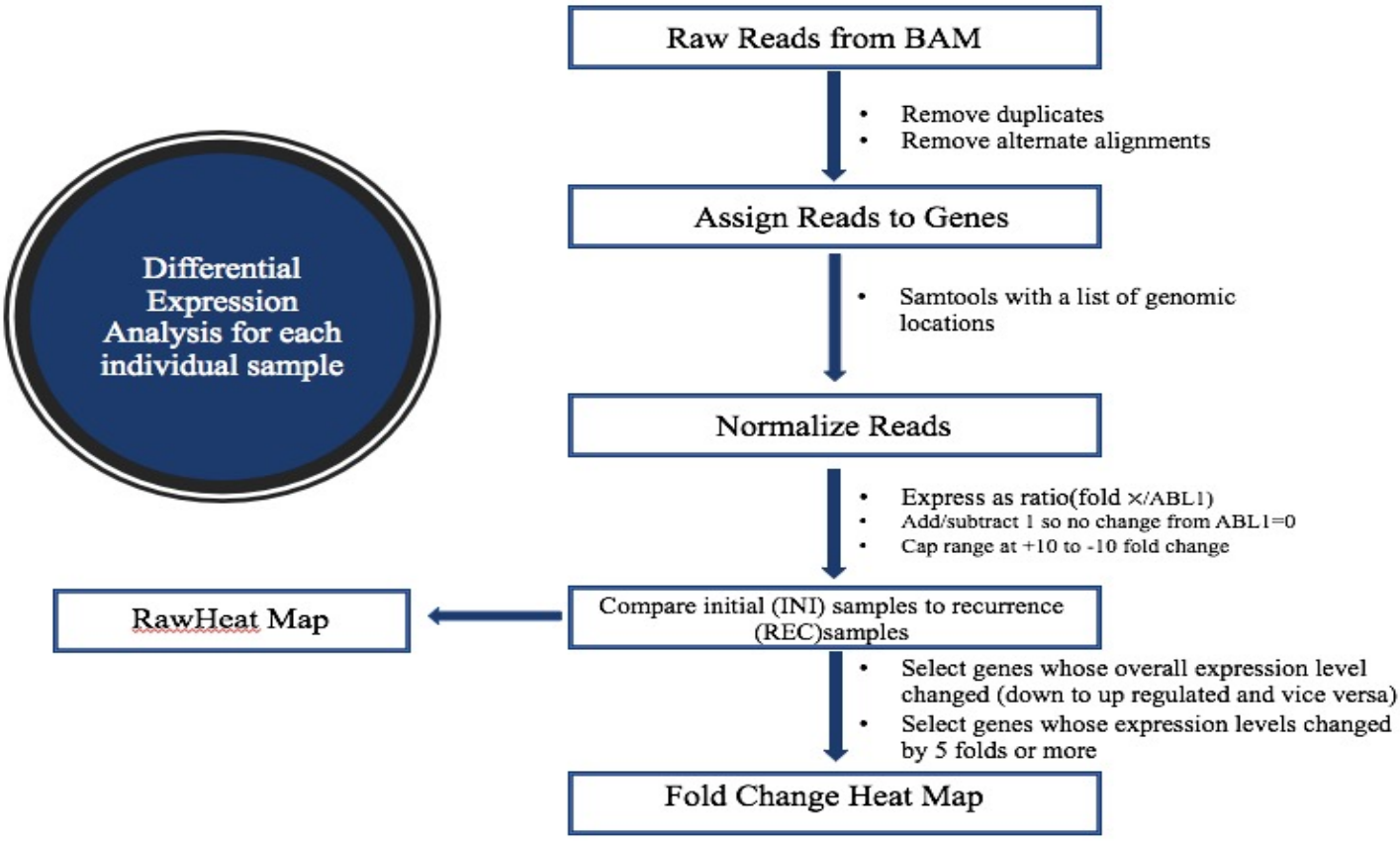

Differential Expression Analysis for each individual sample
Raw Reads from BAM
Remove duplicates
Remove alternate alignments
Assign Reads to Genes
Samtools with a list of genomic locations
Normalize Reads
Express as ratio(fold ×/ABL1)
Add/subtract 1 so no change from ABL1=0
Cap range at +10 to -10 fold change
Compare initial (INI) samples to recurrence (REC)samples
RawHeat Map
Select genes whose overall expression level changed (down to up regulated and vice versa)
Select genes whose expression levels changed by 5 folds or more
Fold Change Heat Map

Figure 4: RNAseq Illumina method evaluation using 2 samples stored as fresh frozen and FFPE. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1 gene. Sample 1 represents the same tumor in both storage conditions at diagnosis, sample 2 represents the same tumor in both storage conditions after therapy relapse. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively).

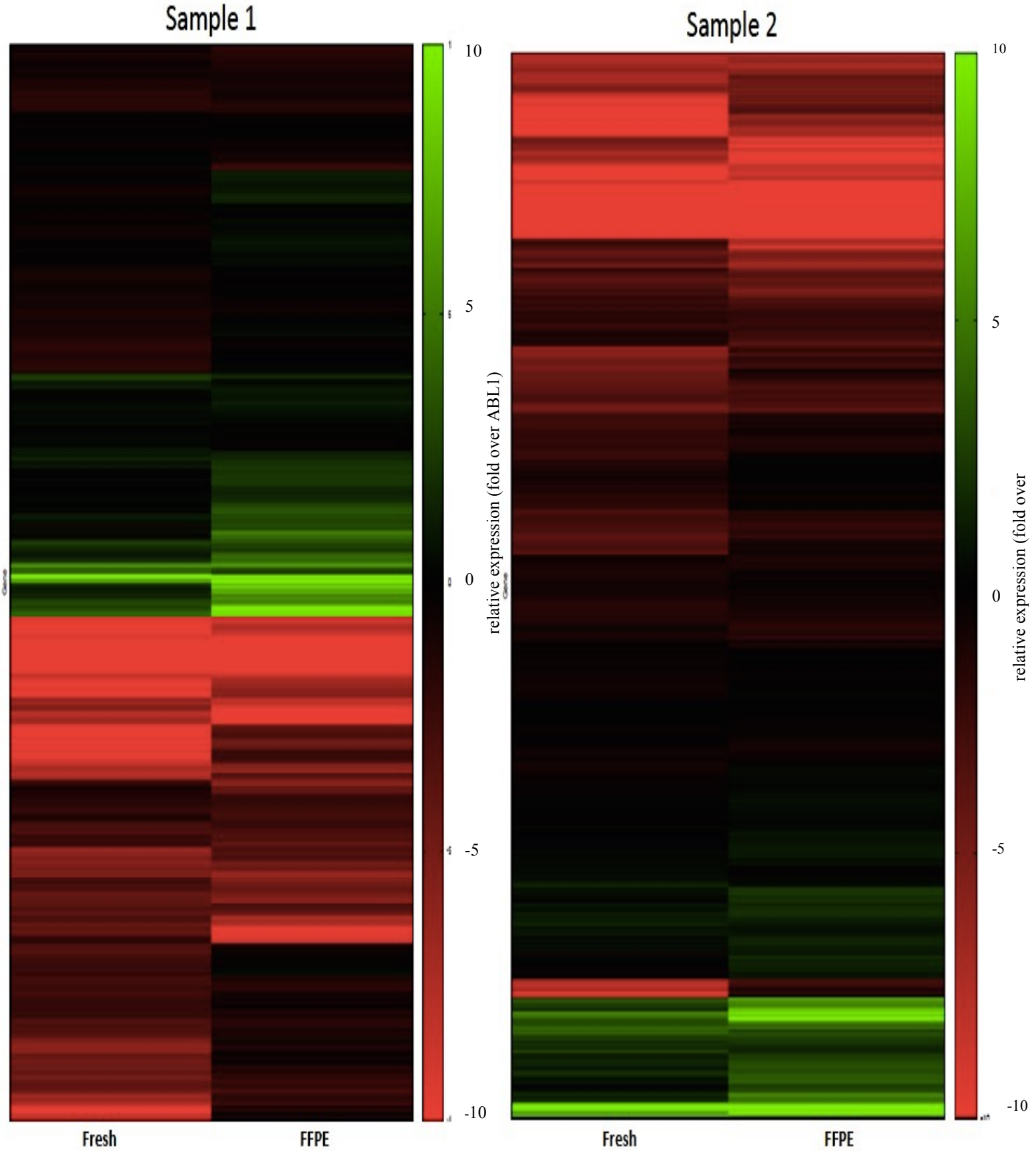
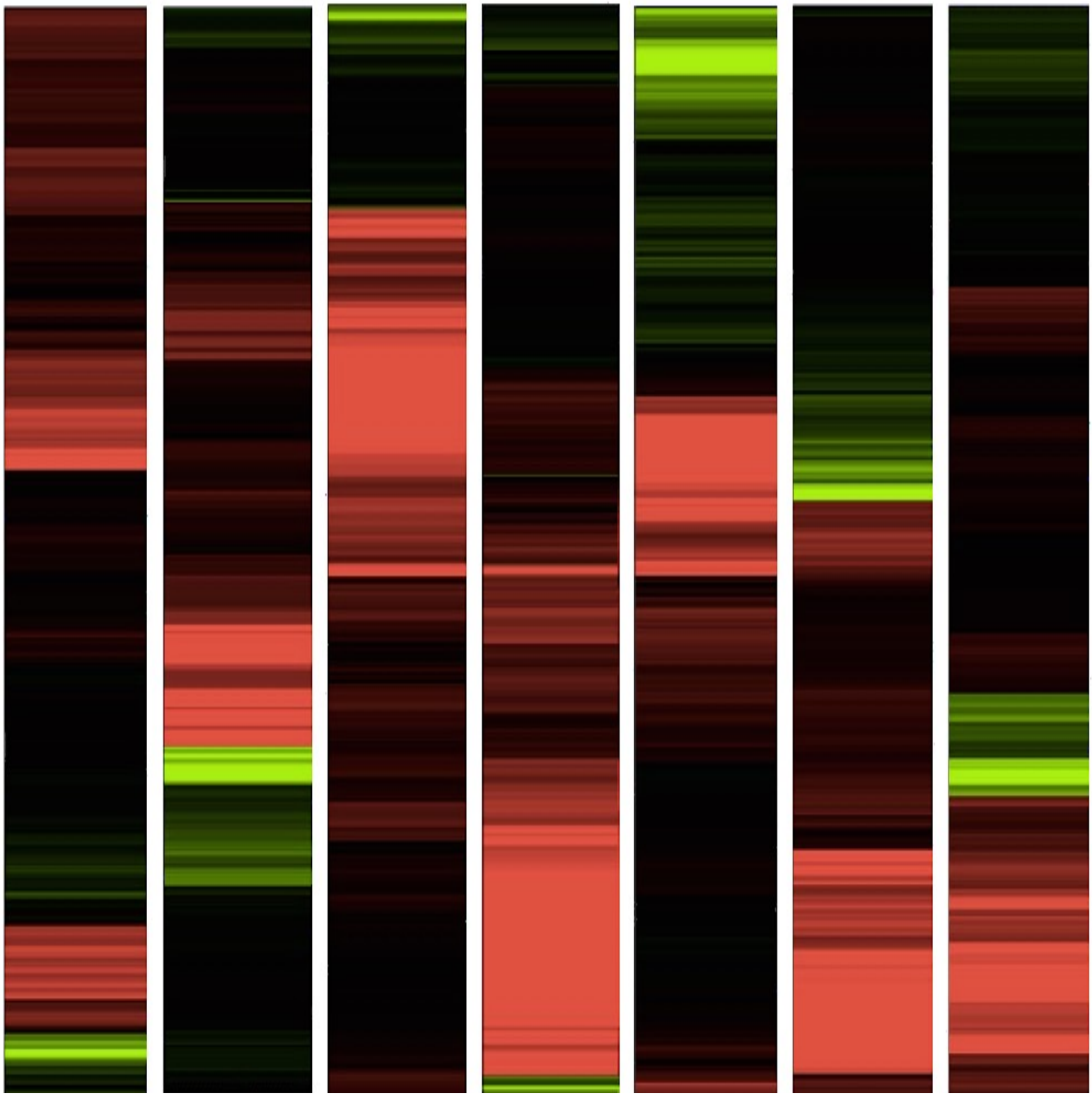

Sample 1
Sample 2
10
5
0
-5
-10
relative expression (fold over ABL1)
relative expression (fold over
Fresh
FFPE
Fresh
FFPE

Figure 5: RNAseq full transcriptome analysis of 7 pre-treatment samples of DLBL. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1.



Sample 1**(1387)** Sample 2**(6100)** Sample 3**(1009)** Sample 4**(2093)** Sample 5**(706)** Sample 6**(13084)** Sample 7**(948)**

Figure 6a: “Cured & fatal/refractory” grouping. Individual cases in this study are represented in columns and genes in rows, with colors representing high (red) or low (black) expression. DLBL samples interpreted as falling into the “cured” group show higher expression of the genes in the upper panel while DLBL samples interpreted as falling into the “fatal/ refractory’ group show higher expression of genes in the lower panel. (Shipp et al. Nature 2002;8;68)

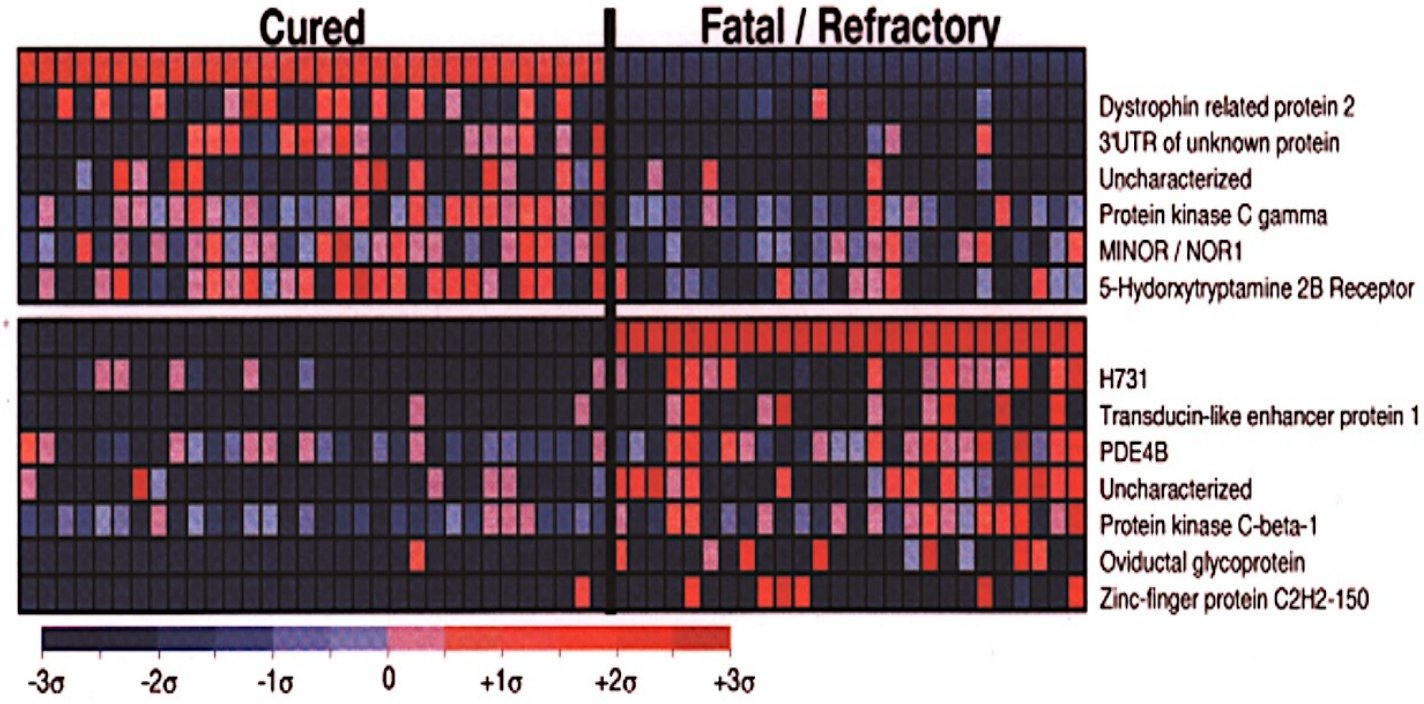

Cured
Fatal / Refractory
Dystrophin related protein 2
3'UTR of unknown protein
Uncharacterized
Protein kinase C gamma
MINOR / NOR1
5-Hydoxytryptamine 2B Receptor
H731
Transducin-like enhancer protein 1
PDE4B
Uncharacterized
Protein kinase C-beta-1
Oviductal glycoprotein
Zinc-finger protein C2H2-150
-3σ
-2σ
-1σ
0
+1σ
+2σ
+3σ

Figure 6b: RNA transcripts from genes listed in figure 6a. The genes represented in the RNAseq data from each pre-therapy at initial diagnosis sample are shown in rows aligned top to bottom as in figure 6a. Each sample would be classified as “fatal/refractory”. Gene names are on the left side. Sample names are at the bottom. Note: there are 8 samples as 1-8948-INI sample represents the fresh frozen storage condition and 1-948 represents the FFPE storage condition for the sample from the same patient at diagnosis. The numbers on the right side represent the number of reads per kb of transcript over ABL1. Colors representing high (red) or low (black) expression.

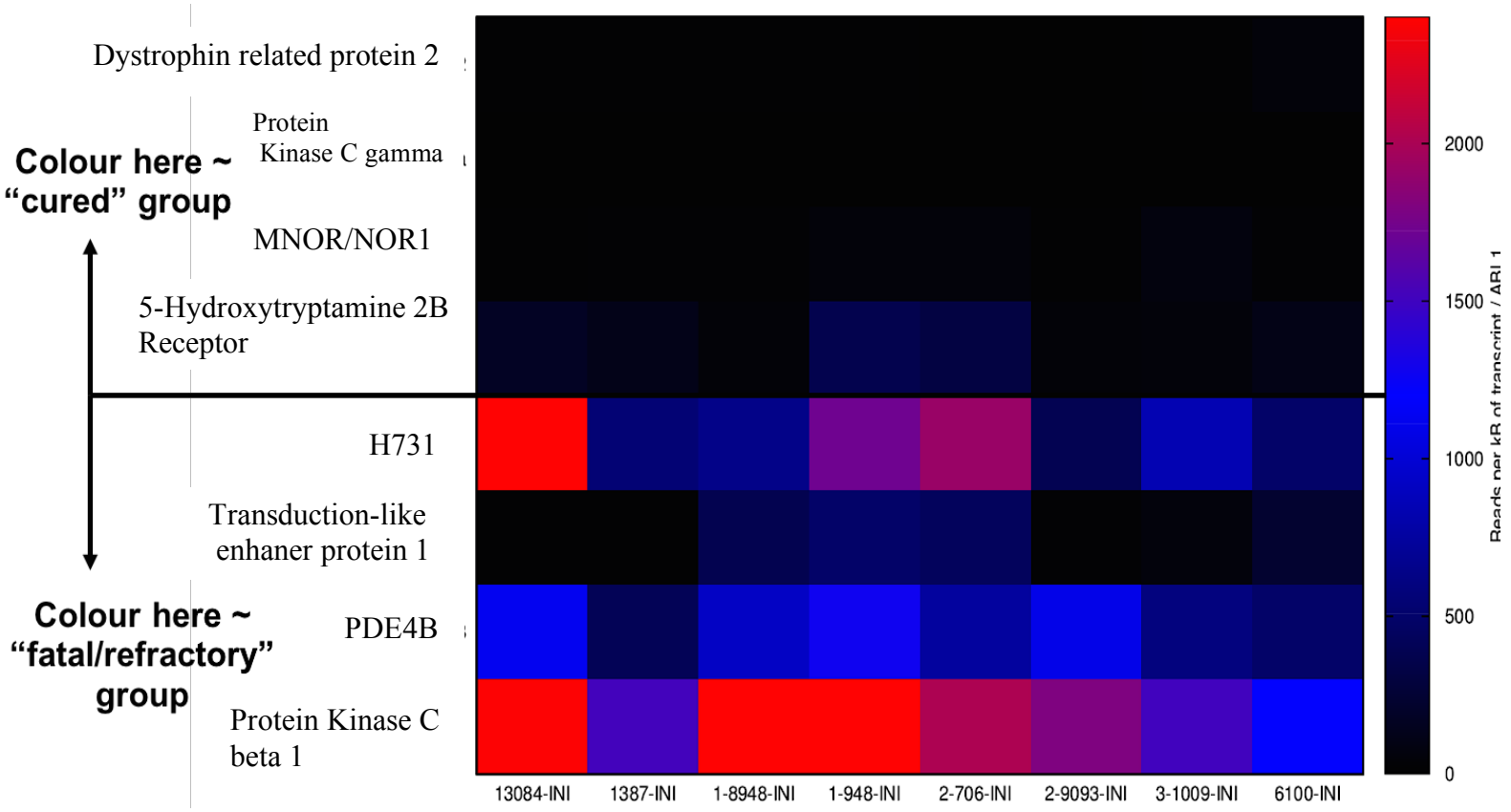

Dystrophin related protein 2
Protein Kinase C gamma
Colour here ~ "cured" group
MNOR/NOR1
5-Hydroxytryptamine 2B Receptor
H731
Transduction-like enhaner protein 1
Colour here ~ "fatal/refractory" group
PDE4B
Protein Kinase C beta 1
13084-INI
1387-INI
1-8948-INI
1-948-INI
2-706-INI
2-9093-INI
3-1009-INI
6100-INI
2000
1500
1000
500
0

Figure 7a: DLBL sample 1 - RNAseq full transcriptome analysis heat map, pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps were generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively).

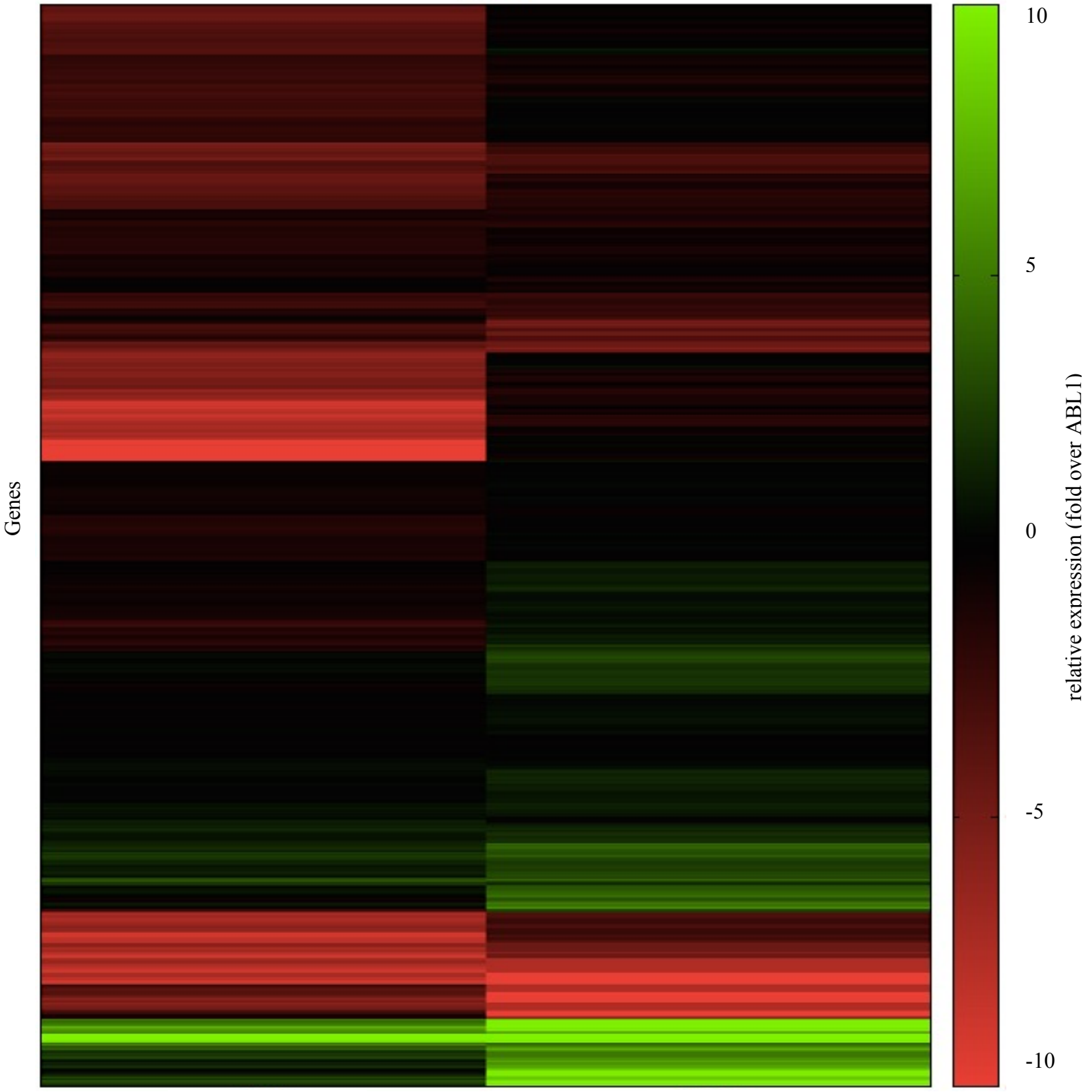

Genes
10
5
0
-5
-10
relative expression (fold over ABL1)
Sample 1- INI
Sample 1- REC

Figure 7b: DLBL sample 2 - RNAseq full transcriptome analysis heatmap, pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively).

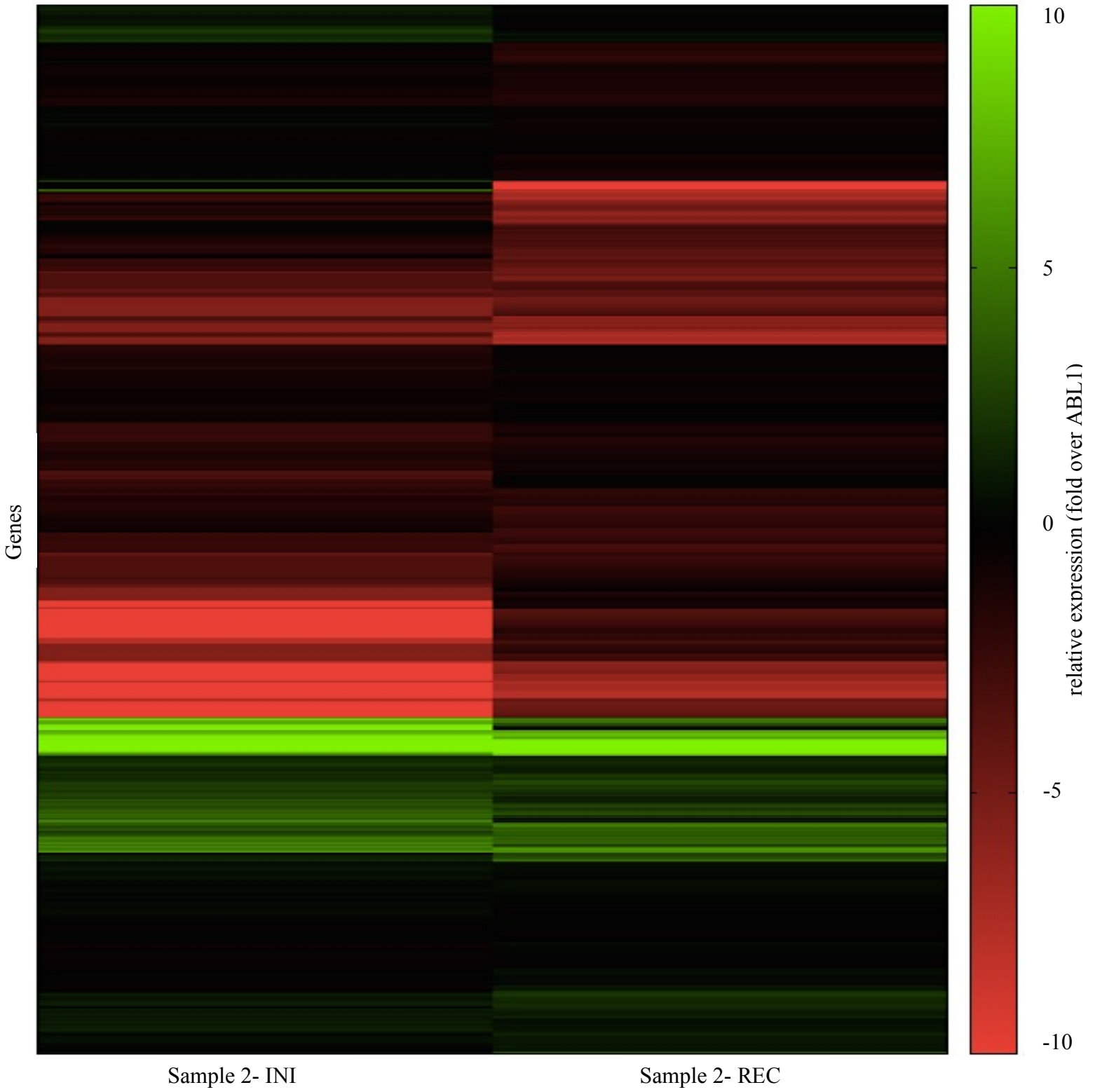

Genes
Sample 2- INI
Sample 2- REC
10
5
0
-5
-10
relative expression (fold over ABL1)

Figure 7c: DLBL sample 3 - RNAseq full transcriptome analysis heatmap, pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively).

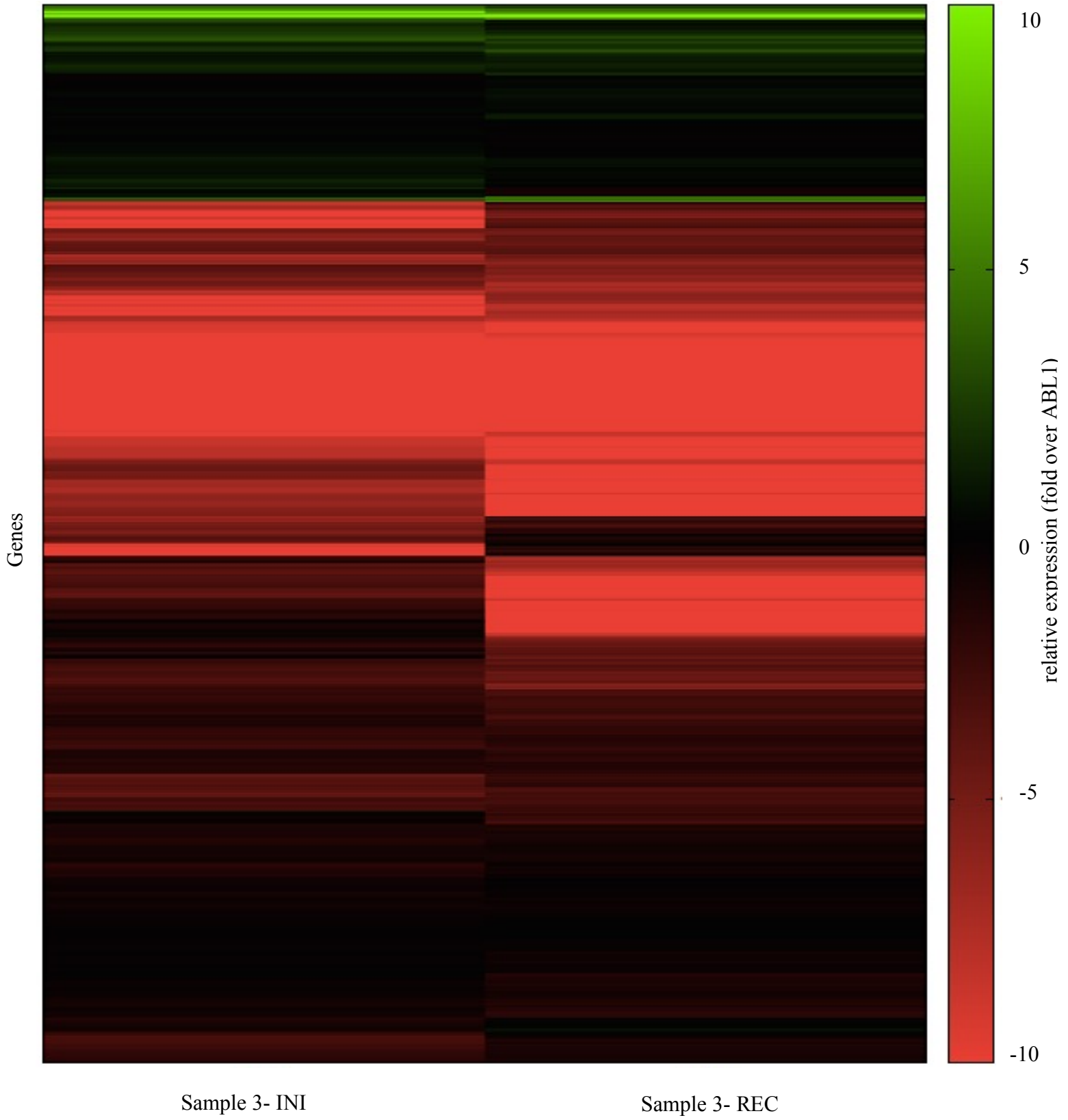

Genes
10
5
0
-5
-10
relative expression (fold over ABL1)
Sample 3- INI
Sample 3- REC

Figure 7d: DLBL sample 4 - RNAseq full transcriptome analysis heatmap, pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively).

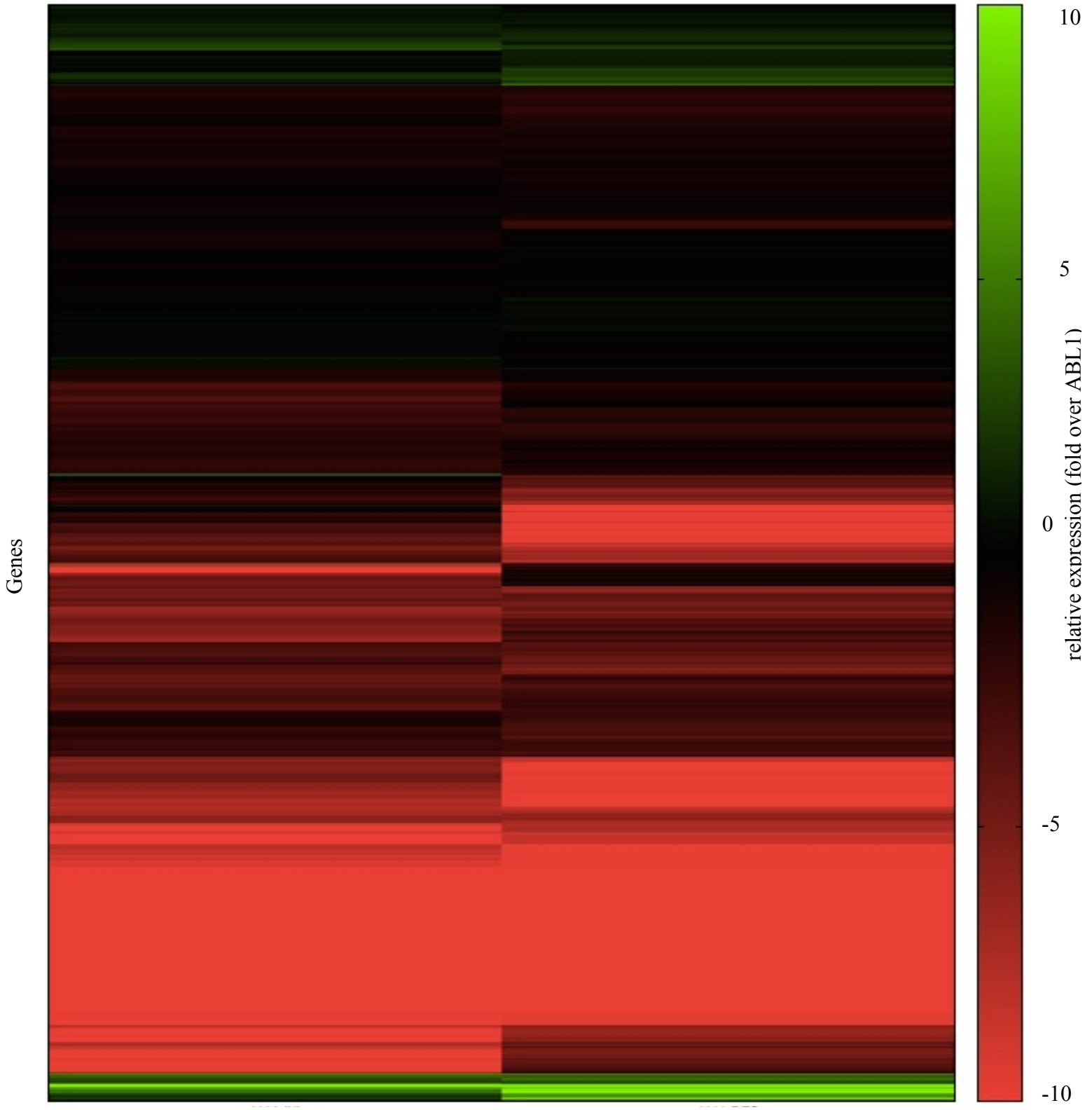

Genes
Sample 4- INI
Sample 4- REC
10
5
0
-5
-10
relative expression (fold over ABL1)

Figure 7e: DLBL sample 5 - RNAseq full transcriptome analysis heatmap, pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively).

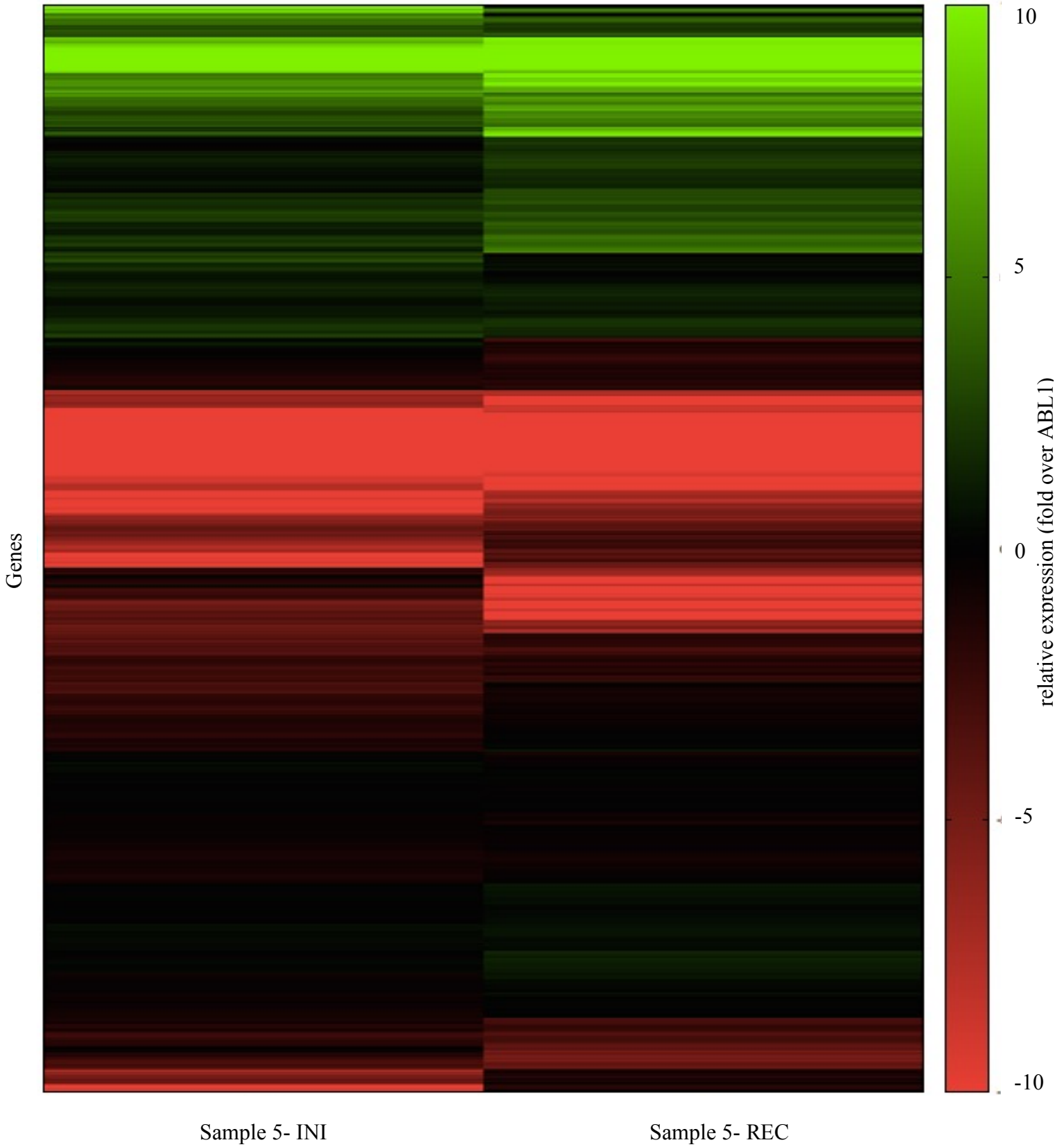

Genes
10
5
0
-5
-10
relative expression (fold over ABL1)
Sample 5- INI
Sample 5- REC

Figure 7f: DLBL sample 6 - RNAseq full transcriptome analysis heat map, pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively).

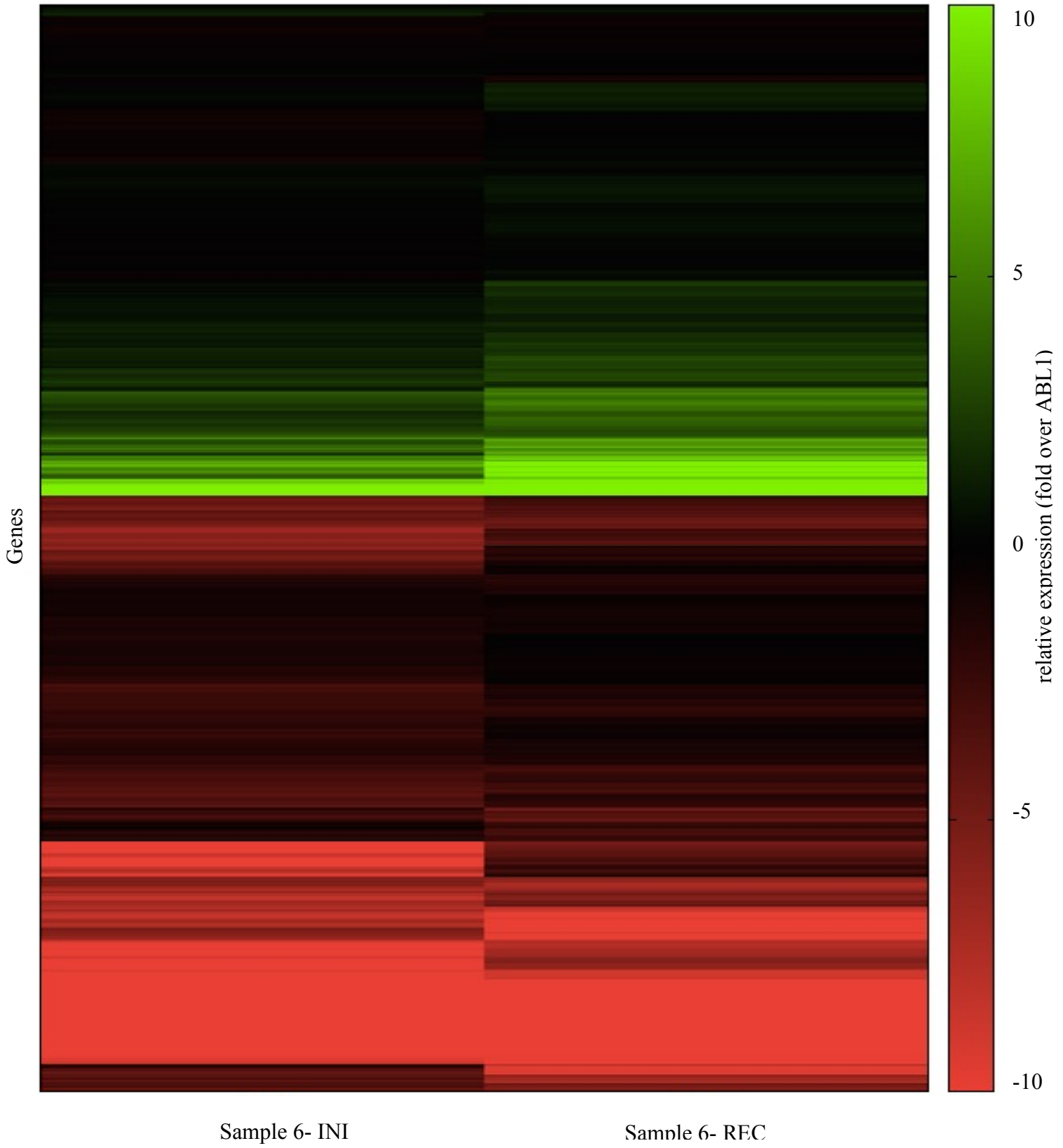

Genes
10
5
0
-5
-10
relative expression (fold over ABL1)
Sample 6- INI
Sample 6- REC

Figure 7g: DLBL sample 7 - RNAseq full transcriptome analysis heat map pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively).

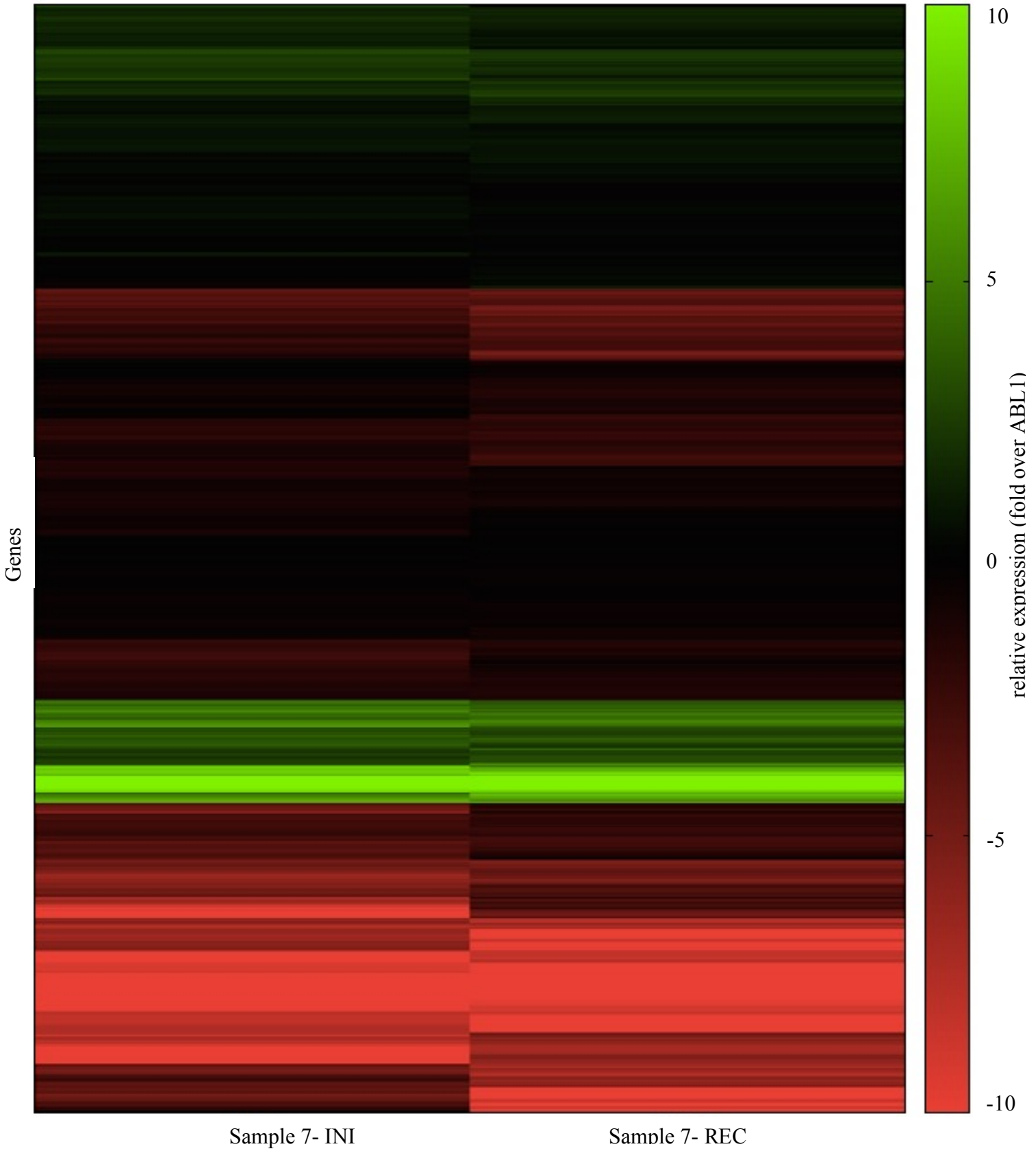

Genes
10
5
0
-5
-10
relative expression (fold over ABL1)
Sample 7- INI
Sample 7- REC

Figure 8a: DLBL sample 1 showing gene expression of genes that changed ≥ 5-fold between pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively). Note: the genes on the left of the figure are also shown in appendix A.

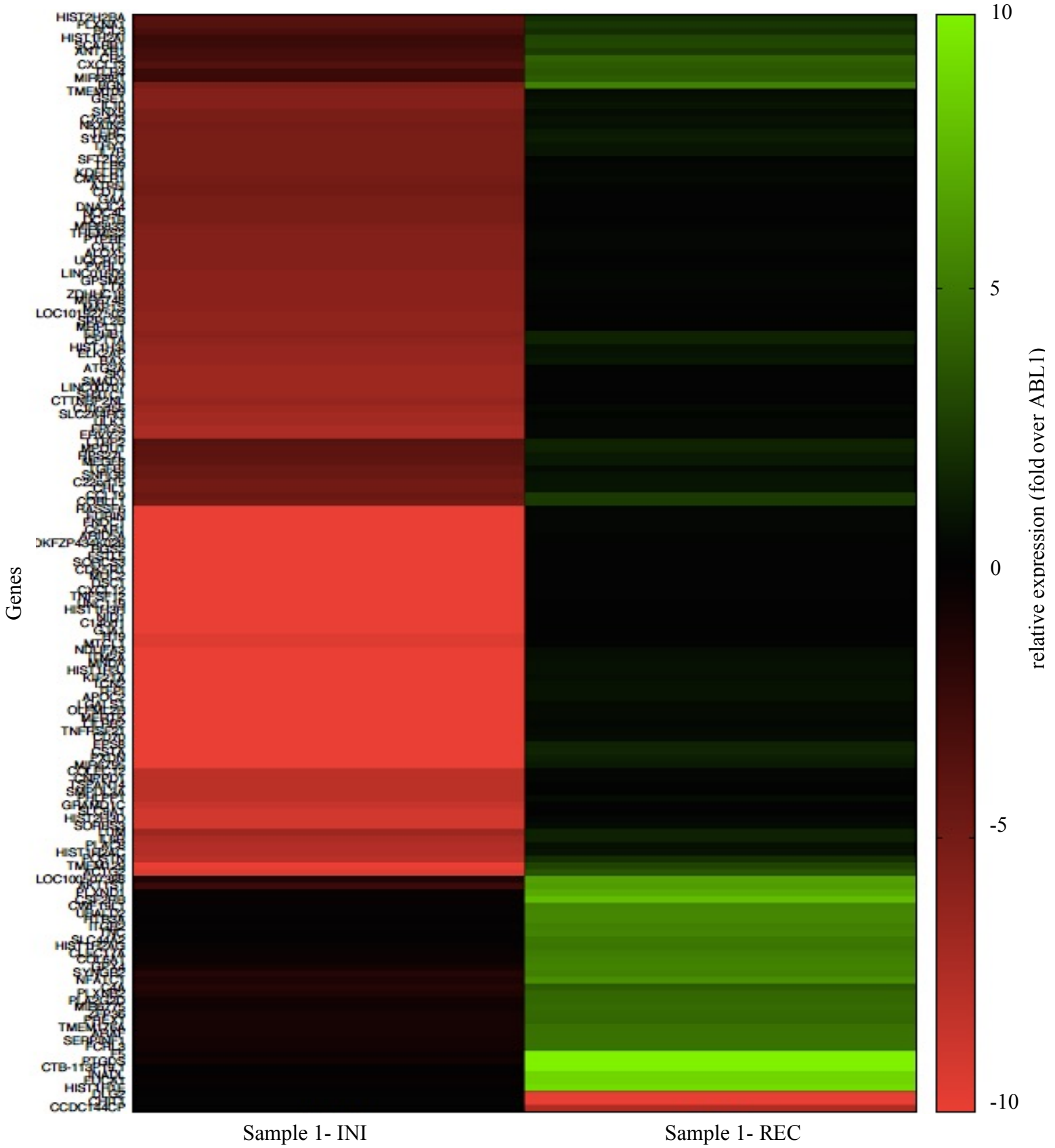

Genes
Sample 1- INI
Sample 1- REC
10
5
0
-5
-10
relative expression (fold over ABL1)

Figure 8b: DLBL sample 2 showing gene expression for genes that changed ≥ 5-fold between pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively). Note: the genes on the left of the figure are also shown in appendix A.

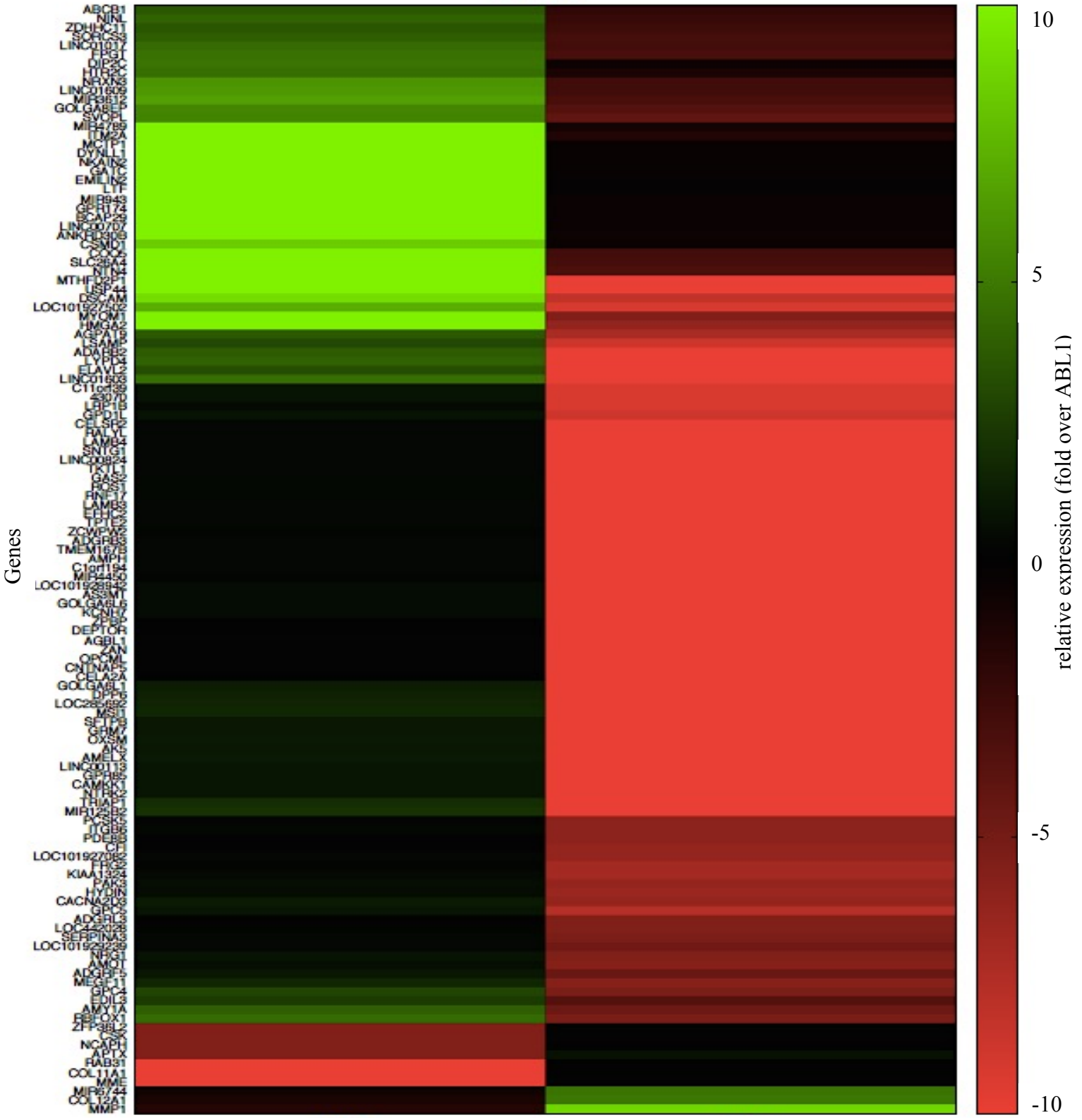

Genes
ABCB1
NINL
ZDHHC11
SORCS3
LINC01017
FPGT
DIP2C
HTR2C
NRXN3
LINC01609
MIR3617
GOLGA8EP
SWAP1
MIR4789
ITM2A
MCTP1
DYNLL1
NKAIN2
GATC
EMILIN2
LTF
MIR943
GPR174
SCAP2
LINC00177
ANKRD30B
CSMD1
COCH
SLC26A4
NTN4
MTHFD2P1
USP44
DSCAM
LOC101927502
MYOM1
HMGA2
ACPP
CSAMP
ADARB2
LYPD4
ELAVL2
LINC01503
C11orf39
S3I7D
LRP1B
GPD1L
CELSR2
RALYL
LAMB4
SNTG1
LINC00824
TKTL1
GAS2
ROS1
RNF17
LAMB3
TEX2
TPTE2
ZCWPW2
ADGRB3
TMEM167B
AMPH
C19orf84
MIR4450
LOC101929842
ASSMT
GOLGA8L6
KCNIP4
PPP2R2B
DEPTOR
AGBL1
ZAN
OPCML
CNTNAP5
CELA2A
GOLGA8L1
DPP6
LOC285692
MSI1
SPTB9
GPRM7
OXSM
AK5
AMELX
LINC00113
GPR85
CAMKK1
NTRK2
THAP1
MIR1253B2
PCSK5
TGFB3
PDE8B
CFI
LOC101927082
FHL2
KIAA1324
PAK3
HYDIN
CACNA2D3
GPC5
ADGRL3
LOC442028
SERPINA3
LOC101928239
NRG1
AMOT
ADGRF5
MEGF11
GPC4
EDIL3
AMY1A
REPOX1
ZFP36L2
CDX2
NCAPH
APTX
RAB31
COL11A1
MME
MIR6744
COL12A1
MMP1
Sample 2- INI
Sample 2- REC
10
5
0
-5
-10
relative expression (fold over ABL1)

Figure 8c: DLBL sample 3 showing gene expression for genes that changed ≥ 5-fold between pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively). Note: the genes on the left of the figure are also shown in appendix A.

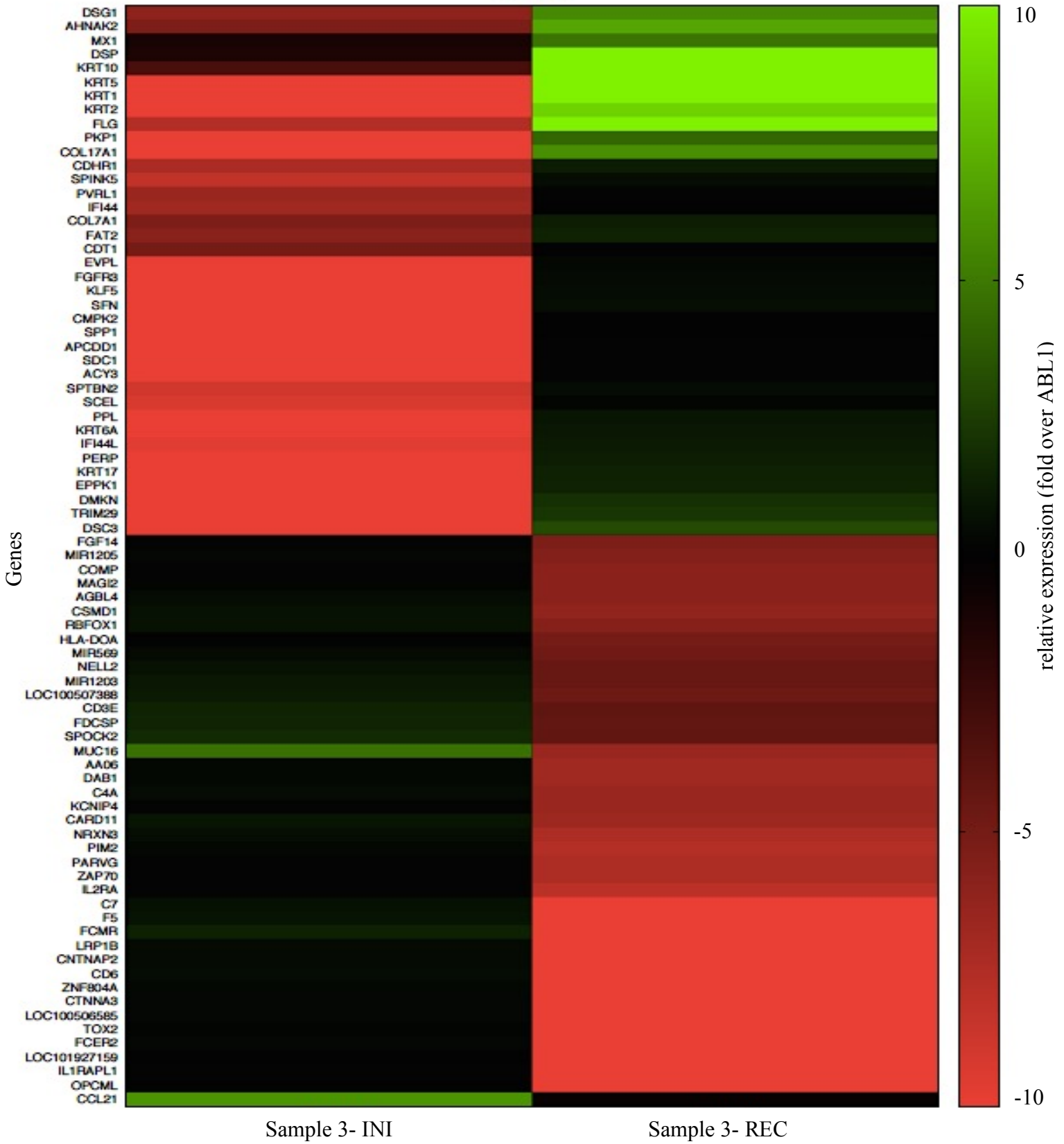

DSG1
AHNAK2
MX1
DSP
KRT10
KRT5
KRT1
KRT2
FLG
PKP1
COL17A1
CDHR1
SPINK5
PVRL1
IFI44
COL7A1
FAT2
CDT1
EVPL
FGFR3
KLF5
SFN
CMPK2
SPP1
APCDD1
SDC1
ACY3
SPTBN2
SCEL
PPL
KRT6A
IFI44L
PERP
KRT17
EPPK1
DMKN
TRIM29
DSC3
FGF14
MIR1205
COMP
MAGI2
AGBL4
CSMD1
RBFOX1
HLA-DOA
MIR569
NELL2
MIR1203
LOC100507388
CD3E
FDCSP
SPOCK2
MUC16
AAD6
DAB1
C4A
KCNIP4
CARD11
NRXN3
PIM2
PARVG
ZAP70
IL2RA
C7
F5
FCMR
LRP1B
CNTNAP2
CD6
ZNF804A
CTNNA3
LOC100506585
TOX2
FCER2
LOC101927159
IL1RAPL1
OPCML
CCL21
Genes
10
5
0
-5
-10
relative expression (fold over ABL1)
Sample 3- INI
Sample 3- REC

Figure 8d: DLBL sample 4 showing gene expression for genes that changed ≥ 5-fold between pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively). Note: the genes on the left of the figure are also shown in appendix A.

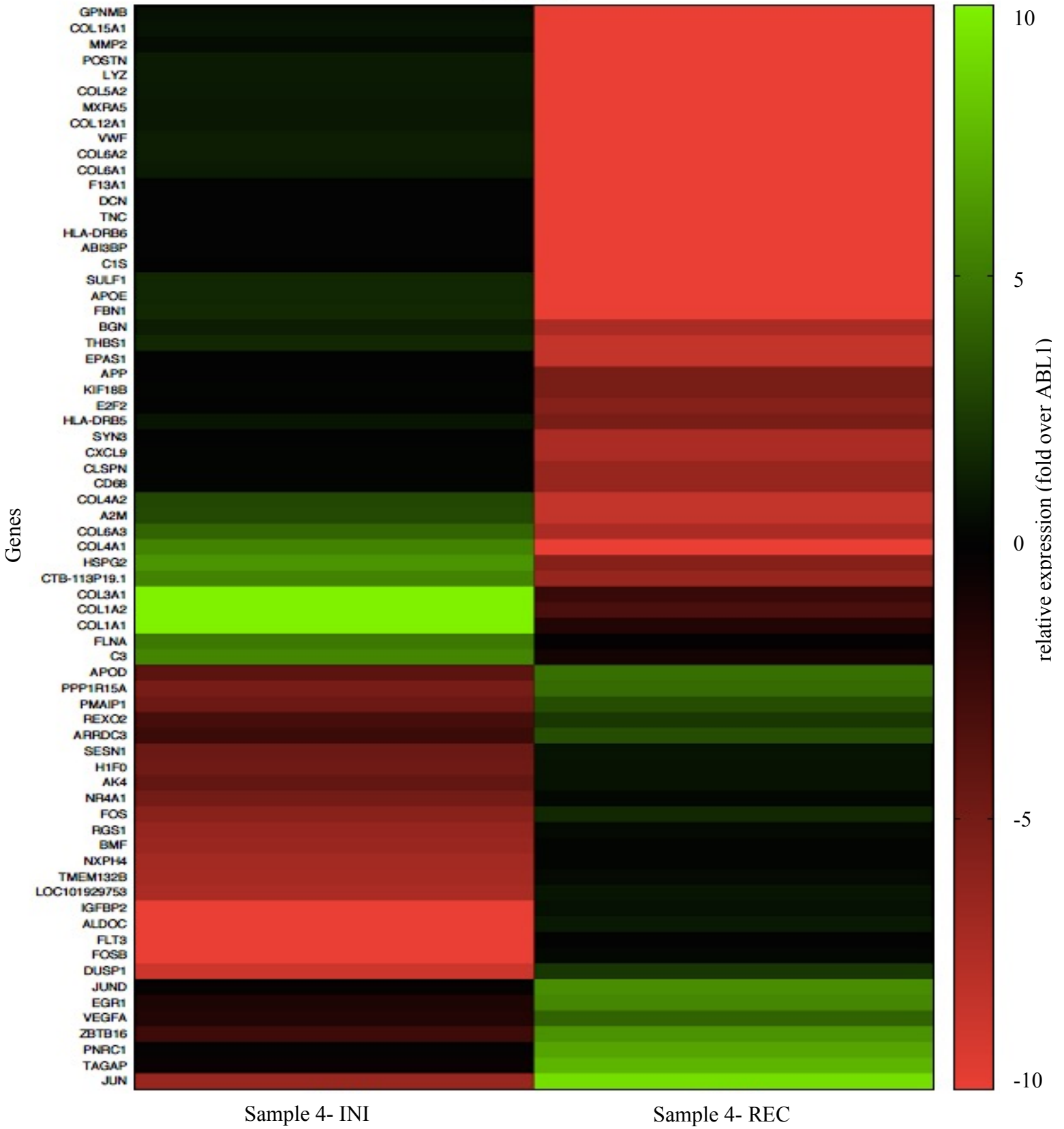

GPNMB
COL15A1
MMP2
POSTN
LYZ
COL5A2
MXRA5
COL12A1
VWF
COL6A2
COL6A1
F13A1
DCN
TNC
HLA-DRB6
ABI3BP
C1S
SULF1
APOE
FBN1
BGN
THBS1
EPAS1
APP
KIF18B
E2F2
HLA-DRB5
SYN3
CXCL9
CLSPN
CD68
COL4A2
A2M
COL6A3
COL4A1
HSPG2
CTB-113P19.1
COL3A1
COL1A2
COL1A1
FLNA
C3
APOD
PPP1R15A
PMAIP1
REXO2
ARRDC3
SESN1
H1F0
AK4
NR4A1
FOS
RGS1
BMF
NXPH4
TMEM132B
LOC101929753
IGFBP2
ALDOC
FLT3
FOSB
DUSP1
JUND
EGR1
VEGFA
ZBTB16
PNRC1
TAGAP
JUN
Genes
Sample 4- INI
Sample 4- REC
10
5
0
-5
-10
relative expression (fold over ABL1)

Figure 8e: DLBL sample 5 showing gene expression for genes that changed ≥ 5-fold between pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively). Note: the genes on the left of the figure are also shown in appendix A.

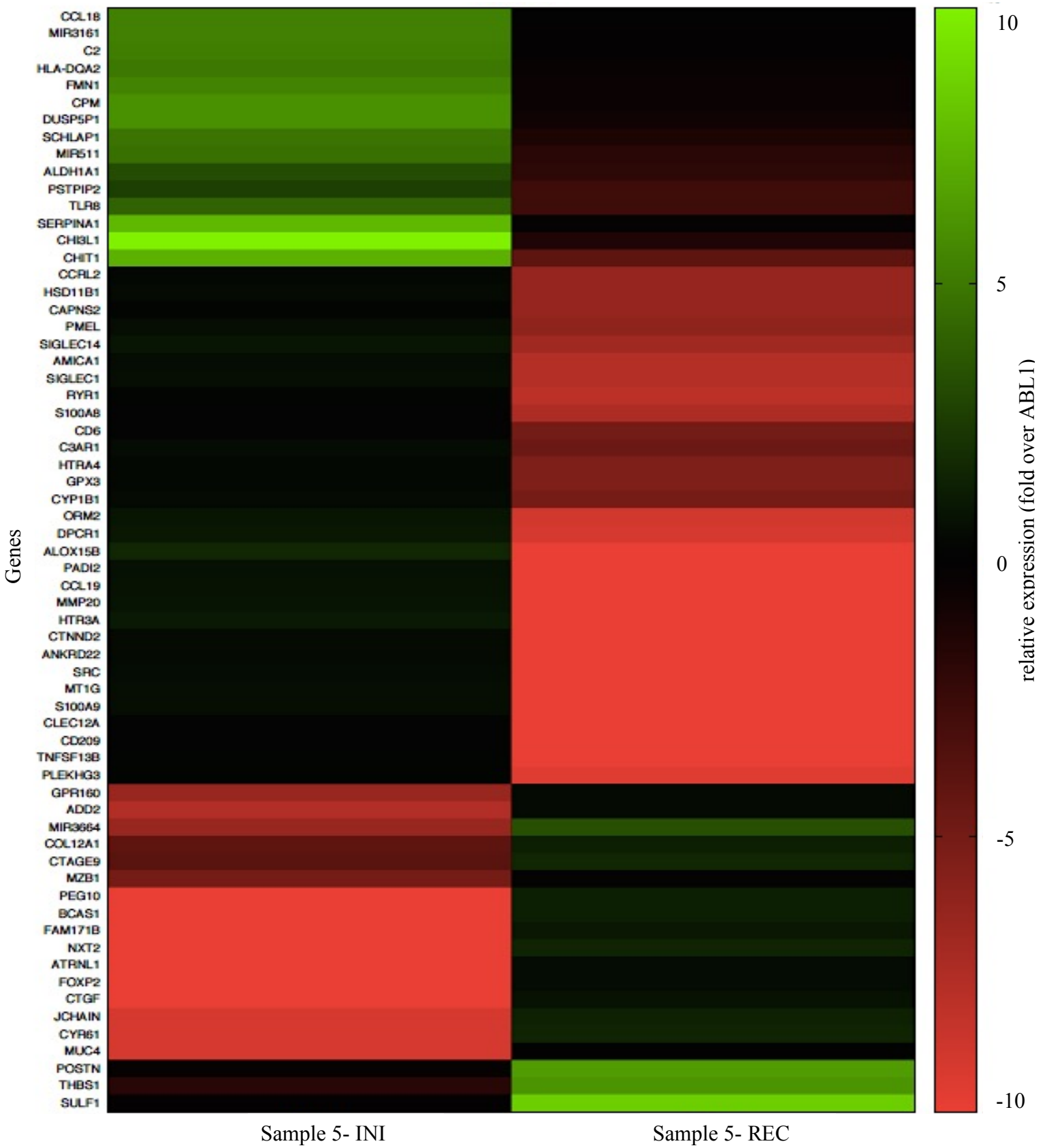

Genes
CCL18
MIR3161
C2
HLA-DQA2
FMN1
CPM
DUSP5P1
SCHLAP1
MIR511
ALDH1A1
PSTPIP2
TLR8
SERPINA1
CHI3L1
CHIT1
CCRL2
HSD11B1
CAPNS2
PMEL
SIGLEC14
AMICA1
SIGLEC1
RYR1
S100A8
CD6
C3AR1
HTRA4
GPX3
CYP1B1
ORM2
DPCR1
ALOX15B
PADI2
CCL19
MMP20
HTR3A
CTNND2
ANKRD22
SRC
MT1G
S100A9
CLEC12A
CD209
TNFSF13B
PLEKHG3
GPR160
ADD2
MIR3664
COL12A1
CTAGE9
MZB1
PEG10
BCAS1
FAM171B
NXT2
ATRNL1
FOXP2
CTGF
JCHAIN
CYR61
MUC4
POSTN
THBS1
SULF1
Sample 5- INI
Sample 5- REC
10
5
0
-5
-10
relative expression (fold over ABL1)

Figure 8f: DLBL sample 6 showing gene expression for genes that changed ≥ 5-fold between pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1.The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively). Note: the genes on the left of the figure are also shown in appendix A.

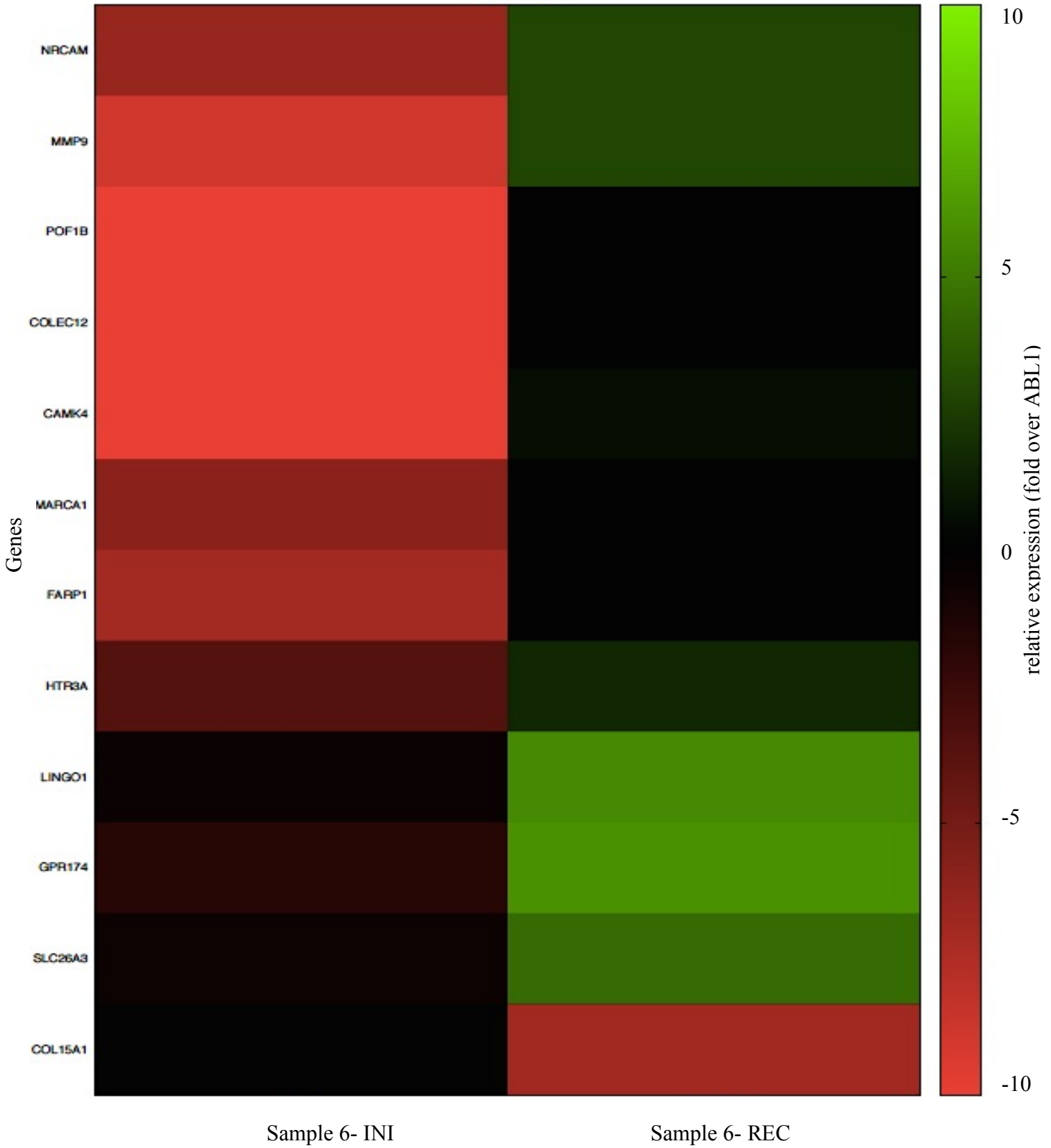

Genes
NRCAM
MMP9
POF1B
COLEC12
CAMK4
MARCA1
FARP1
HTR3A
LINGO1
GPR174
SLC26A3
COL15A1
Sample 6- INI
Sample 6- REC
10
5
0
-5
-10
relative expression (fold over ABL1)

Figure 8-g: DLBL sample 7 showing gene expression for genes that changed ≥ 5-fold between pre-therapy at initial diagnosis sample (left) and after therapy relapse sample (right). After aligning the genes to the reference transcriptome, then normalization, heatmaps generated using gnuplot 5.0. patchlevel 6. Green color represents upregulated genes, red color represents downregulated genes, black no change in relative to ABL1. The numbers on the right-hand side of the graph represent normalized gene expression (fold change in relative to ABL1). Values above 10 are also possible, the legend is to cap the color scale (any values above 10 or below -10 are assigned the same color as 10 and -10 respectively). Note: the genes on the left of the figure are also shown in appendix A.

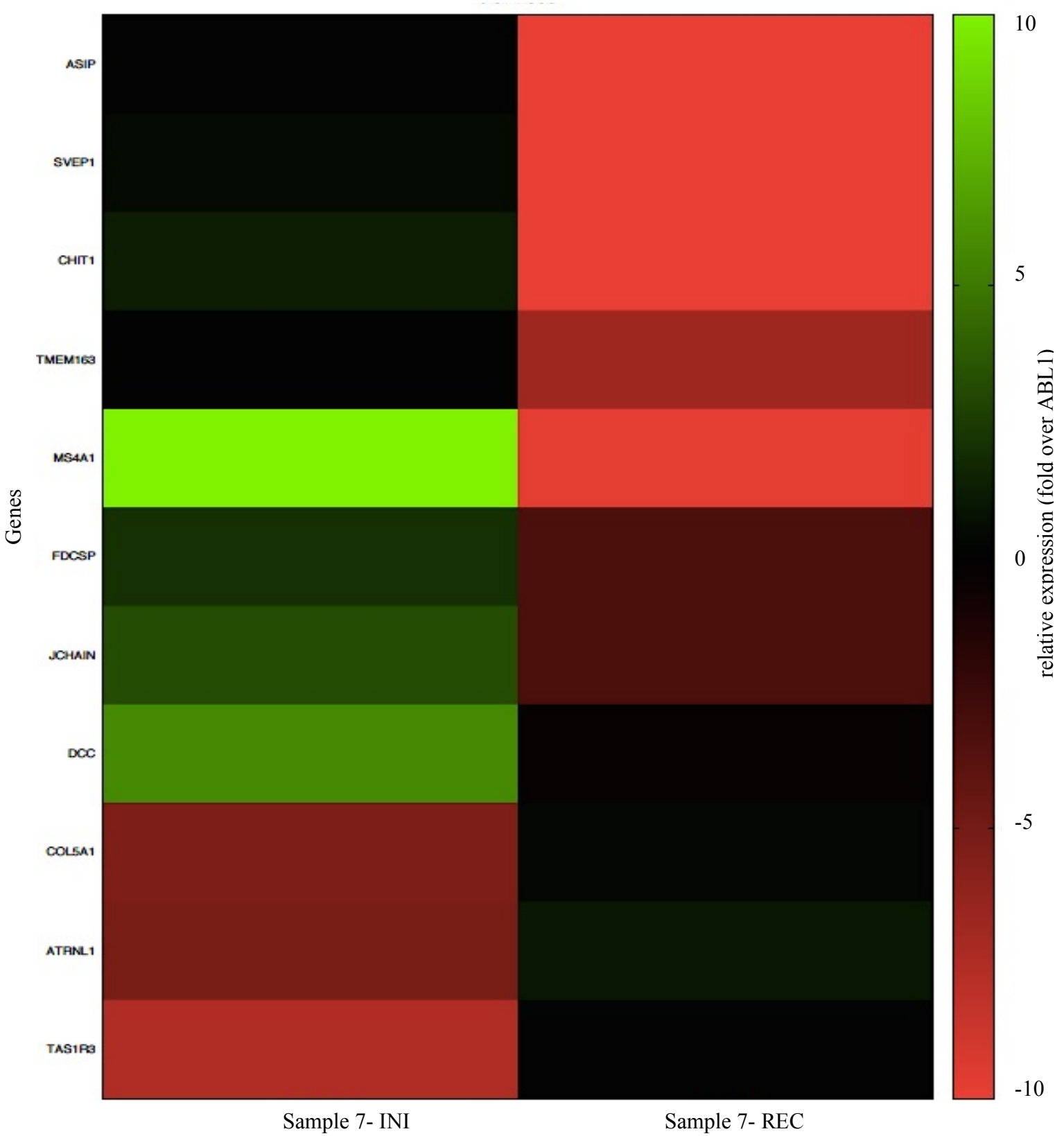

ASIP
SVEP1
CHIT1
TMEM163
MS4A1
FDCSP
JCHAIN
DCC
COL5A1
ATRNL1
TAS1R3
Genes
Sample 7- INI
Sample 7- REC
10
5
0
-5
-10
relative expression (fold over ABL1)

Figure 9a: DLBL sample 1 – Pathway analysis screenshots. Top DAVID's tool showing the altered pathways and the counts (numbers) of altered genes in each pathway (see Table 5, which shows the altered genes in each pathway). Bottom Panther tool showing the biological functions of altered genes in the sample. Note: Blue stars on the left side in DAVID's tool screen shot indicates that the same pathway is altered in another sample as well.

| | Term | RT | Genes | Count | % | P-Value |
|---|---|---|---|---|---|---|
| | Systemic lupus erythematosus | RT | | 9 | 0.0 | 8.3E-5 |
| | Cytokine-cytokine receptor interaction | RT | | 11 | 0.1 | 1.5E-4 |
| | Alcoholism | RT | | 8 | 0.0 | 2.7E-3 |
| ★ | Complement and coagulation cascades | RT | | 5 | 0.0 | 6.2E-3 |
| | Malaria | RT | | 4 | 0.0 | 1.5E-2 |
| | Staphylococcus aureus infection | RT | | 4 | 0.0 | 2.0E-2 |
| | Pertussis | RT | | 4 | 0.0 | 4.6E-2 |
| ★ | NF-kappa B signaling pathway | RT | | 4 | 0.0 | 6.6E-2 |

Displaying only results with P<0.05; click here to display all results

| | Homo sapiens (REF) | Client Text Box Input (▿ Hierarchy NEW! ⓘ) | | | | |
|---|---|---|---|---|---|---|
| PANTHER GO-Slim Biological Process | # | # | expected | Fold Enrichment | +/- | P value |
| regulation of gene expression, epigenetic | 51 | 7 | .37 | 18.81 | + | 3.01E-05 |
| chromatin organization | 263 | 11 | 1.92 | 5.73 | + | 1.06E-03 |
| ↳organelle organization | 752 | 19 | 5.49 | 3.46 | + | 6.82E-04 |
| ↳cellular component organization | 1584 | 31 | 11.56 | 2.68 | + | 9.87E-05 |
| ↳cellular component organization or biogenesis | 1724 | 31 | 12.58 | 2.46 | + | 5.87E-04 |
| response to external stimulus | 311 | 10 | 2.27 | 4.41 | + | 2.60E-02 |
| cell adhesion | 481 | 14 | 3.51 | 3.99 | + | 3.25E-03 |
| ↳biological adhesion | 481 | 14 | 3.51 | 3.99 | + | 3.25E-03 |
| nervous system development | 668 | 15 | 4.87 | 3.08 | + | 2.99E-02 |
| Unclassified | 8633 | 48 | 62.98 | .76 | - | 0.00E00 |

Table 5: Sample 1 altered genes in the pathways identified by DAVID tool, and their expression in the recurrence tumor.

| Pathway | genes | expression in recurrence |
|---|---|---|
| Cytokine-cytokine receptor interaction | CCL19 | High |
| | CXCL12 | High |
| | CXCL13 | High |
| | CD70 | High |
| | CSF2RB | High |
| | CSF2RB | High |
| | IL10 | High |
| | IL6R | High |
| | IL7R | High |
| | LTA | High |
| | TNFSF12 | High |
| Pathway | genes | expression in recurrence |
| Complement and coagulation cascades | F5 | High |
| | CR2 | High |
| | C4A | High |
| | CSAR1 | High |
| | TFPI | High |
| Pathway | genes | expression in recurrence |
| NF-kappa B signaling pathway | CCL19 | High |
| | CXCL12 | High |
| | LTA | High |
| | TLR4 | High |

Figure 9b: DLBL sample 2– Pathway analysis screenshots. Top DAVID's tool showing the altered pathways and the counts (numbers) of altered genes in each pathway (see Table 6, which shows the altered genes in each pathway). Bottom Panther tool showing the biological functions of altered genes in the sample. Note: Blue stars on the left side in DAVID's tool screen shot indicates that the same pathway is altered in another sample as well.

| Term | RT | Genes | Count | % | P-Value |
|---|---|---|---|---|---|
| ECM-receptor interaction | RT | | 4 | 0.0 | 1.1E-2 |
| Protein digestion and absorption | RT | | 4 | 0.0 | 1.1E-2 |
| Focal adhesion | RT | | 5 | 0.0 | 2.3E-2 |

Displaying only results with P<0.05; click here to display all results

| | Homo sapiens (REF) | Client Text Box Input (∇ Hierarchy NEW!) | | | | |
|---|---|---|---|---|---|---|
| PANTHER GO-Slim Biological Process | # | # | expected | Fold Enrichment | +/- | P value |
| nervous system development | 668 | 12 | 3.34 | 3.59 | + | 3.28E-02 |
| Unclassified | 8633 | 47 | 43.22 | 1.09 | + | 0.00E00 |

Table 6: Sample 2, altered genes in the pathways identified by DAVID tool, and their expression in the recurrence tumor.

| Pathway | genes | expression in recurrence |
|---|---|---|
| Extracellular matrix--receptor interaction | COL11A1 | High |
| | ITGB6 | Low |
| | LAMB3 | Low |
| | LAMB4 | Low |
| Pathway | genes | expression in recurrence |
| Focal adhesion | COL11A1 | High |
| | ITGB6 | Low |
| | LAMB3 | Low |
| | LAMB4 | Low |
| | PAK3 | Low |

Figure 9c: DLBL sample 3– Pathway analysis screenshots. Top DAVID's tool showing the altered pathways and the counts (numbers) of altered genes in each pathway (see Table 7, which shows the altered genes in each pathway). Bottom Panther tool showing the biological functions of altered genes in the sample. Note: Blue stars on the left side in DAVID's tool screen shot indicates that the same pathway is altered in another sample as well.

| Sublist | Category | Term | RT | Genes | Count | % | P-Value |
|---|---|---|---|---|---|---|---|
| ☐ | KEGG_PATHWAY | Cell adhesion molecules (CAMs) | RT | | 5 | 0.1 | 2.7E-3 |
| ☐ | KEGG_PATHWAY | Staphylococcus aureus infection | RT | | 3 | 0.0 | 2.1E-2 |
| ☐ | KEGG_PATHWAY | Complement and coagulation cascades | RT | | 3 | 0.0 | 3.4E-2 |
| ☐ | KEGG_PATHWAY | Hematopoietic cell lineage | RT | | 3 | 0.0 | 4.9E-2 |
| ☐ | KEGG_PATHWAY | NF-kappa B signaling pathway | RT | | 3 | 0.0 | 5.1E-2 |
| ☐ | KEGG_PATHWAY | ECM-receptor interaction | RT | | 3 | 0.0 | 5.1E-2 |
| ☐ | KEGG_PATHWAY | PI3K-Akt signaling pathway | RT | | 5 | 0.1 | 5.4E-2 |
| ☐ | KEGG_PATHWAY | T cell receptor signaling pathway | RT | | 3 | 0.0 | 6.9E-2 |

Displaying only results with P<0.05; click here to display all results

| | Homo sapiens (REF) | Client Text Box Input (▽ Hierarchy NEW! ⓘ) | | | | |
|---|---|---|---|---|---|---|
| PANTHER GO-Slim Biological Process | # | # | expected | Fold Enrichment | +/- | P value |
| response to external stimulus | 311 | 7 | 1.11 | 6.29 | + | 3.12E-02 |
| cellular component morphogenesis | 545 | 9 | 1.95 | 4.62 | + | 3.48E-02 |
| cellular process | 8199 | 45 | 29.32 | 1.53 | + | 4.84E-02 |
| Unclassified | 8633 | 24 | 30.87 | .78 | - | 0.00E00 |

Table 7: Sample 3, altered genes in the pathways identified by DAVID tool, and their expression in the recurrence tumor.

| Pathway | genes | expression in recurrence |
|---|---|---|
| Cell adhesion molecules (CAMs) | CD6 | Low |
| | CNTNAP2 | Low |
| | HLA-DOA | Low |
| | NRXN3 | Low |
| | SDC1 | High |
| Pathway | genes | expression in recurrence |
| Complement and coagulation cascades | F5 | Low |
| | C4A | Low |
| | C7 | Low |
| Pathway | genes | expression in recurrence |
| Hematopoietic cell lineage | CD3E | Low |
| | FCER2 | Low |
| | IL2RA | Low |
| Pathway | genes | expression in recurrence |
| Extracellular matrix-receptor interaction | COMP | Low |
| | SPP1 | High |
| | SD1 | High |
| Pathway | genes | expression in recurrence |
| NF-kappa B signaling pathway | CCL21 | Low |
| | CARD11 | Low |
| | ZAP70 | Low |
| Pathway | genes | expression in recurrence |
| PI3K-Akt signaling pathway | COMP | Low |
| | FGF14 | Low |
| | FGFR3 | Low |
| | IL2RA | low |
| | SPP1 | High |
| Pathway | genes | expression in recurrence |
| T-cell receptor signaling pathway | CD3E | Low |
| | CARD11 | Low |
| | ZAP70 | Low |

Figure 9d: DLBL sample 4– Pathway analysis screenshots. Top DAVID's tool showing the altered pathways and the counts (numbers) of altered genes in each pathway (see Table 8, which shows the altered genes in each pathway). Bottom Panther tool showing the biological functions of altered genes in the sample. Note: Blue stars on the left side in DAVID's tool screen shot indicates that the same pathway is altered in another sample as well.

| Category | Term | RT | Genes | Count | % |
|---|---|---|---|---|---|
| KEGG_PATHWAY | ECM-receptor interaction | RT | | 13 | 0.1 |
| KEGG_PATHWAY | Focal adhesion | RT | | 15 | 0.1 |
| KEGG_PATHWAY | Protein digestion and absorption | RT | | 11 | 0.1 |
| KEGG_PATHWAY | PI3K-Akt signaling pathway | RT | | 14 | 0.1 |
| KEGG_PATHWAY | Amoebiasis | RT | | 6 | 0.1 |
| KEGG_PATHWAY | Pathways in cancer | RT | | 10 | 0.1 |
| KEGG_PATHWAY | Complement and coagulation cascades | RT | | 5 | 0.0 |
| KEGG_PATHWAY | Bladder cancer | RT | | 4 | 0.0 |
| KEGG_PATHWAY | Proteoglycans in cancer | RT | | 6 | 0.1 |
| KEGG_PATHWAY | Platelet activation | RT | | 5 | 0.0 |
| KEGG_PATHWAY | Leishmaniasis | RT | | 4 | 0.0 |
| KEGG_PATHWAY | Pertussis | RT | | 4 | 0.0 |
| KEGG_PATHWAY | Rheumatoid arthritis | RT | | 4 | 0.0 |
| KEGG_PATHWAY | MAPK signaling pathway | RT | | 6 | 0.1 |
| KEGG_PATHWAY | Staphylococcus aureus infection | RT | | 3 | 0.0 |
| KEGG_PATHWAY | Osteoclast differentiation | RT | | 4 | 0.0 |
| KEGG_PATHWAY | Renal cell carcinoma | RT | | 3 | 0.0 |
| KEGG_PATHWAY | Amphetamine addiction | RT | | 3 | 0.0 |
| KEGG_PATHWAY | Hepatitis B | RT | | 4 | 0.0 |
| KEGG_PATHWAY | p53 signaling pathway | RT | | 3 | 0.0 |
| KEGG_PATHWAY | HTLV-I infection | RT | | 5 | 0.0 |

Displaying only results with P<0.05; click here to display all results

| | Homo sapiens (REF) | Client Text Box Input (▽ Hierarchy NEW! ⓘ) | | | | |
|---|---|---|---|---|---|---|
| PANTHER GO-Slim Biological Process | # | # | expected | Fold Enrichment | +/- | P value |
| cell-cell adhesion | 305 | 7 | .96 | 7.29 | + | 1.22E-02 |
| ↳cell adhesion | 481 | 8 | 1.51 | 5.28 | + | 3.24E-02 |
| ↳biological adhesion | 481 | 8 | 1.51 | 5.28 | + | 3.24E-02 |
| immune system process | 1269 | 14 | 3.99 | 3.51 | + | 8.20E-03 |
| Unclassified | 8633 | 22 | 27.17 | .81 | - | 0.00E00 |

Table 8: Sample 4, altered genes in the pathways identified by DAVID tool, and their expression in the recurrence tumor.

| Pathway | genes | expression in recurrence | Pathway | genes | expression in recurrence |
|---|---|---|---|---|---|
| Extracellular matrix--receptor interaction | COL1A1 | Low | Pathways in cancer | E2F2 | Low |
| | COL1A2 | Low | | FOS | High |
| | COL3A1 | Low | | JUN | High |
| | COL4A1 | Low | | COL4A1 | Low |
| | COL4A2 | Low | | COL4A2 | Low |
| | COL5A2 | Low | | EPAS1 | Low |
| | COL6A1 | Low | | FLT3 | High |
| | COL6A2 | Low | | MMP2 | Low |
| | COL6A3 | Low | | VEGFA | High |
| | HSPG2 | Low | | ZBTB16 | High |
| | TNC | Low | Pathway | genes | expression in recurrence |
| | THBS1 | Low | Complement and coagulation cascades | A2M | Low |
| | VWF | Low | | F13A1 | Low |
| Pathway | genes | expression in recurrence | | C1S | Low |
| Focal adhesion | JUN | High | | C3 | Low |
| | COL1A1 | Low | | VWF | Low |
| | COL1A2 | Low | Pathway | genes | expression in recurrence |
| | COL3A1 | Low | Bladder cancer | E2F2 | Low |
| | COL4A1 | Low | | MMp2 | Low |
| | COL4A2 | Low | | THBS1 | Low |
| | COL5A2 | Low | | VEGFA | High |
| | COL6A1 | Low | Pathway | genes | expression in recurrence |
| | COL6A2 | Low | Proteoglycans in cancer | DCN | Low |
| | COL6A3 | Low | | FLNA | Low |
| | FLNA | Low | | HSPG2 | Low |
| | TNC | Low | | MMP2 | Low |
| | THBS1 | Low | | THBS1 | Low |
| | VWF | Low | | VEGFA | High |

| Pathway | genes | expression in recurrence | Pathway | genes | expression in recurrence |
|---|---|---|---|---|---|
| PI3K-Akt signaling pathway | COL1A1 | Low | Platelet activation | COL1A1 | Low |
| | COL1A2 | Low | | COL1A2 | Low |
| | COL3A1 | Low | | COL3A1 | Low |
| | COL4A1 | Low | | COL5A2 | Low |
| | COL4A2 | Low | | VWF | Low |
| | COL5A2 | Low | Pathway | genes | expression in recurrence |
| | COL6A1 | Low | MAPK signaling pathway | FOS | High |
| | COL6A2 | Low | | JUN | High |
| | COL6A3 | Low | | JUND | High |
| | NR4A1 | Low | | DUSP1 | High |
| | TNCL | Low | | FLNA | Low |
| | THBS1 | Low | | NR4A1 | High |
| | VEGFA | Low | Pathway | genes | expression in recurrence |
| | VWF | Low | Renal cell carcinoma | JUN | High |
| Pathway | genes | expression in recurrence | | EPAS1 | Low |
| p53 signaling pathway | PMAIP | High | | VEGFA | High |
| | SESN1 | High | | | |
| | THBS1 | Low | | | |

Figure 9e: DLBL sample 5– Pathway analysis screenshots. Top DAVID's tool showing the altered pathways and the counts (numbers) of altered genes in each pathway (see Table 9, which shows the altered genes in each pathway). Bottom Panther tool showing the biological functions of altered genes in the sample. Note: Blue stars on the left side in DAVID's tool screen shot indicates that the same pathway is altered in another sample as well.

| Category | Term | RT | Genes | Count | % | P-Value |
|---|---|---|---|---|---|---|
| KEGG_PATHWAY | Staphylococcus aureus infection | RT | | 3 | 0.0 | 1.5E-2 |
| KEGG_PATHWAY | Complement and coagulation cascades | RT | | 3 | 0.0 | 2.4E-2 |
| KEGG_PATHWAY | Cell adhesion molecules (CAMs) | RT | | 3 | 0.0 | 8.6E-2 |
| KEGG_PATHWAY | Phagosome | RT | | 3 | 0.0 | 9.8E-2 |

Displaying only results with P<0.05; click here to display all results

| | Homo sapiens (REF) | Client Text Box Input (▽ Hierarchy NEW! ?) | | | | |
|---|---|---|---|---|---|---|
| PANTHER GO-Slim Biological Process | # | # | expected | Fold Enrichment | +/- | P value |
| cell adhesion | 481 | 8 | 1.35 | 5.91 | + | 1.44E-02 |
| ↳biological adhesion | 481 | 8 | 1.35 | 5.91 | + | 1.44E-02 |
| immune system process | 1269 | 14 | 3.57 | 3.92 | + | 2.10E-03 |
| Unclassified | 8633 | 21 | 24.29 | .86 | - | 0.00E00 |

Table 9: Sample 5, altered genes in the pathways identified by DAVID tool, and their expression in the recurrence tumor.

| Pathway | genes | expression in recurrence |
|---|---|---|
| Complement and coagulation cascades | C2 | Low |
| | C3AR1 | Low |
| | SERPINA1 | Low |
| Pathway | genes | expression in recurrence |
| Cell adhesion molecules (CAMs) | CD6 | Low |
| | HLA-DQA2 | Low |
| | SIGLEC1 | Low |
| Pathway | genes | expression in recurrence |
| Phagosome | CD209 | Low |
| | HLA-DQA2 | High |
| | THBS1 | High |

# References


1. Spear, Brian B., Margo Heath-Chiozzi, and Jeffrey Huff. "Clinical application of pharmacogenetics." Trends in molecular medicine 7.5 (2001): 201-204.
2. Sehn, Laurie H., et al. "Introduction of combined CHOP plus rituximab therapy dramatically improved outcome of diffuse large B-cell lymphoma in British Columbia." *Journal of Clinical Oncology* 23.22 (2005): 5027-5033.
3. Shipp, Margaret A., et al. "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning." *Nature medicine* 8.1 (2002): 68-74.
4. Lohr, Jens G., et al. "Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing." *Proceedings of the National Academy of Sciences* 109.10 (2012): 3879-3884.
5. Sesques, Pierre, and Nathalie A. Johnson. "Approach to the diagnosis and treatment of high-grade B-cell lymphomas with MYC and BCL2 and/or BCL6 rearrangements." *Blood* 129.3 (2017): 280-288.
6. Alizadeh, Ash A., et al. "Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling." *Nature*403.6769 (2000): 503-511.
7. Shipp, Margaret A., et al. "Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning." *Nature medicine* 8.1 (2002): 68-74.
8. Beillard, E., et al. "Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using ‘real-time’quantitative reverse-transcriptase

polymerase chain reaction (RQ-PCR)–a Europe against cancer program." *Leukemia* 17.12 (2003): 2474-2486.

9. Goubran, Hadi A., et al. "Regulation of tumor growth and metastasis: the role of tumor microenvironment." *Cancer growth and metastasis* 7 (2014): 9.
10. Scott, David W., and Randy D. Gascoyne. "The tumour microenvironment in B cell lymphomas." *Nature Reviews Cancer* 14.8 (2014): 517-534.
11. Dranoff, Glenn. "Cytokines in cancer pathogenesis and cancer therapy." *Nature Reviews Cancer* 4.1 (2004): 11-22.
12. Damås, J. K., et al. "Homeostatic chemokines CCL19 and CCL21 promote inflammation in human immunodeficiency virus-infected patients with ongoing viral replication." *Clinical & Experimental Immunology* 157.3 (2009): 400-407.
13. Peng, Cheng, et al. "The effect of CCL19/CCR7 on the proliferation and migration of cell in prostate cancer." *Tumor Biology* 36.1 (2015): 329-335.
14. Cassier, Philippe A., et al. "Prognostic value of the expression of C-Chemokine Receptor 6 and 7 and their ligands in non-metastatic breast cancer." *BMC cancer* 11.1 (2011): 213.
15. Hwang, Hyundoo, et al. "Human breast cancer-derived soluble factors facilitate CCL19-induced chemotaxis of human dendritic cells." *Scientific reports* 6 (2016).
16. Wang, Lingyan, et al. "The role of SDF-1/CXCR4 in the vasculogenesis and remodeling of cerebral arteriovenous malformation." *Therapeutics and clinical risk management* 11 (2015): 1337.
17. Guo, F., et al. "CXCL12/CXCR4: a symbiotic bridge linking cancer cells and their stromal neighbors in oncogenic communication networks." *Oncogene* 35.7 (2016): 816-826.

18. Widney, Daniel P., et al. "Expression and function of the chemokine, CXCL13, and its receptor, CXCR5, in Aids-associated non-Hodgkin's lymphoma." *AIDS research and treatment* 2010 (2010).
19. Chen, Lujia, et al. "The expression of CXCL13 and its relation to unfavorable clinical characteristics in young breast cancer." *Journal of translational medicine* 13.1 (2015): 168.
20. Singh, Shailesh, et al. "Serum CXCL13 positively correlates with prostatic disease, prostate-specific antigen and mediates prostate cancer cell invasion, integrin clustering and cell adhesion." *Cancer letters* 283.1 (2009): 29-35.
21. Qi, X. W., et al. "Expression features of CXCR5 and its ligand, CXCL13 associated with poor prognosis of advanced colorectal cancer." *Eur Rev Med Pharmacol Sci* 18.13 (2014): 1916-1924.
22. Rubenstein, James L., et al. "CXCL13 plus interleukin 10 is highly specific for the diagnosis of CNS lymphoma." *Blood*121.23 (2013): 4740-4748.
23. Van De Ven, Koen, and Jannie Borst. "Targeting the T-cell co-stimulatory CD27/CD70 pathway in cancer immunotherapy: rationale and potential." (2015).
24. Jacobs, J., et al. "CD70: An emerging target in cancer immunotherapy." *Pharmacology & therapeutics* 155 (2015): 1-10.
25. Benschop, Robert, Tao Wei, and Songqing Na. "Tumor necrosis factor receptor superfamily member 21: TNFR-related death receptor-6, DR6." *Therapeutic Targets of the TNF Superfamily* (2009): 186-194.
26. Yang, X., et al. "Death receptor 6 (DR6) is required for mouse B16 tumor angiogenesis via the NF-κB, P38 MAPK and STAT3 pathways." *Oncogenesis* 5.3 (2016): e206.

27. Yang, Kun, et al. "DR6 as a diagnostic and predictive biomarker in adult sarcoma." *PloS one* 7.5 (2012): e36525.

28. Saraiva, Margarida, and Anne O'garra. "The regulation of IL-10 production by immune cells." *Nature reviews immunology*10.3 (2010): 170-181.

29. Mannino, Mark H., et al. "The paradoxical role of il-10 in immunity and cancer." *Cancer letters* 367.2 (2015): 103-107.

30. Hong, David S., Laura S. Angelo, and Razelle Kurzrock. "Interleukin-6 and its receptor in cancer." *Cancer* 110.9 (2007): 1911-1928.

31. Kumar, Janani, and Alister C. Ward. "Role of the interleukin 6 receptor family in epithelial ovarian cancer and its clinical implications." *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer* 1845.2 (2014): 117-125.

32. Zenatti, P.P., et al., *Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia.* Nature genetics, 2011

33. Lin, Jack, et al. "The role of IL-7 in Immunity and Cancer." *Anticancer Research* 37.3 (2017): 963-967.

34. Calmon-Hamaty, Flavia, et al. "Lymphotoxin α revisited: general features and implications in rheumatoid arthritis." *Arthritis research & therapy* 13.4 (2011): 232.

35. Morosetti, Roberta, et al. "TWEAK in inclusion-body myositis muscle: possible pathogenic role of a cytokine inhibiting myogenesis." *The American journal of pathology* 180.4 (2012): 1603-1613.

36. Michaelson, Jennifer S., and Linda C. Burkly. "Therapeutic targeting of TWEAK/Fn14 in cancer: exploiting the intrinsic tumor cell killing capacity of the pathway." *Death Receptors and Cognate Ligands in Cancer*. Springer Berlin Heidelberg, 2009. 145-160.

37. Klimatcheva, Ekaterina, et al. "CXCL13 antibody for the treatment of autoimmune disorders." *BMC immunology* 16.1 (2015): 6.

38. Jacobs, J., et al. "CD70: An emerging target in cancer immunotherapy." *Pharmacology & therapeutics* 155 (2015): 1-10.

39. Llorente, Luis, et al. "Clinical and biologic effects of anti–interleukin-10 monoclonal antibody administration in systemic lupus erythematosus." *Arthritis & Rheumatology* 43.8 (2000): 1790-1800.

40. Hunter, Christopher A., and Simon A. Jones. "IL-6 as a keystone cytokine in health and disease." *Nature immunology*16.5 (2015): 448-457.

41. Cheng, Emily, et al. "TWEAK/Fn14 axis-targeted therapeutics: moving basic science discoveries to the clinic." *Frontiers in immunology* 4 (2013).

42. Li, Peipei, et al. "Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data." *BMC bioinformatics* 16.1 (2015): 347.

# Appendix A

Genes ≥ 5-fold differentially expressed genes in the recurrence sample compared to the initial DLBL sample 1. (H in REC: High in recurrence), (L in REC: Low in recurrence). Actual data are shown Figure 8a.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ACTG2 | H in REC | DCP1B | H in REC | LGALS1 | H in REC | RGS2 | H in REC |
| AKT1S1 | H in REC | DKFZP434K028 | H in REC | LILRB2 | H in REC | RPS27L | H in REC |
| ALOX5 | H in REC | DLG2 | L in REC | LINC00707 | H in REC | SCARB1 | H in REC |
| ANTXR1 | H in REC | DNAJC4 | H in REC | LINC01609 | H in REC | SERPINF1 | H in REC |
| APOC2 | H in REC | DSC1 | H in REC | LOC100507388 | H in REC | SFT2D2 | H in REC |
| ARAF | H in REC | ELK2AP | H in REC | LOC101927502 | H in REC | SH3TC1 | H in REC |
| ARID5A | H in REC | EPHB1 | H in REC | LTA | H in REC | SKI | H in REC |
| ATG2A | H in REC | EPS8 | H in REC | LTBP2 | H in REC | SLC2A4RG | H in REC |
| ATP5I | H in REC | ERVV-2 | H in REC | LUM | H in REC | SLC44A2 | H in REC |
| BAX | H in REC | F5 | H in REC | MAP1S | H in REC | SLC9A1 | H in REC |
| BCL3 | H in REC | FCRL3 | H in REC | MEGF8 | H in REC | SMAD1 | H in REC |
| BGN | H in REC | FNDC1 | H in REC | MERTK | H in REC | SMPDL3A | H in REC |
| C10orf55 | L in REC | FPGS | H in REC | MIR6748 | H in REC | SNHG8 | H in REC |
| C14orf1 | H in REC | FSTL5 | H in REC | MIR6775 | H in REC | SNX9 | H in REC |
| C22orf15 | H in REC | FUCA1 | H in REC | MIR6795 | H in REC | SORBS3 | H in REC |
| C4A | H in REC | FURIN | H in REC | MIR6833 | H in REC | SORCS3 | H in REC |
| C5AR1 | H in REC | GAA | H in REC | MIR6881 | H in REC | SPPL2B | H in REC |
| C7orf73 | H in REC | GJA1 | H in REC | MNDA | H in REC | SYNGR2 | H in REC |
| CCDC144CP | L in REC | GPSM3 | H in REC | MPDU1 | H in REC | SYNPO | H in REC |
| CCL19 | H in REC | GPX4 | H in REC | MRPL11 | H in REC | TCN2 | H in REC |
| CD70 | H in REC | GRAMD1C | H in REC | MTCL1 | H in REC | TERC | H in REC |
| CDK5R1 | H in REC | GSE1 | H in REC | MUC2 | H in REC | TFPI | H in REC |
| CDT1 | H in REC | H19 | H in REC | NDUFA3 | H in REC | TGFBI | H in REC |
| CETP | H in REC | HIST1H1E | H in REC | NFATC1 | H in REC | THEMIS2 | H in REC |
| CHIT1 | L in REC | HIST1H2AC | H in REC | NID1 | H in REC | THY1 | H in REC |
| CHL1 | H in REC | HIST1H2AG | H in REC | NKAIN2 | H in REC | TLR4 | H in REC |
| CLEC17A | H in REC | HIST1H2AI | H in REC | NOC4L | H in REC | TLR9 | H in REC |
| CMKLR1 | H in REC | HIST1H3H | H in REC | OLFML2B | H in REC | TMEM109 | H in REC |
| CNPPD1 | H in REC | HIST1H3I | H in REC | PHLPP1 | H in REC | TMEM129 | H in REC |
| COBLL1 | H in REC | HIST1H3J | H in REC | PLA2G2D | H in REC | TMEM176A | H in REC |
| COL6A1 | H in REC | HIST2H2BA | H in REC | PLAC8 | H in REC | TNC | H in REC |
| COLEC12 | H in REC | HIST2H3D | H in REC | PLXNA1 | H in REC | TNFRSF21 | H in REC |
| CPT1A | H in REC | HTR3A | H in REC | PLXNB2 | H in REC | TNFSF12 | H in REC |
| CR2 | H in REC | IL10 | H in REC | PLXND1 | H in REC | TSPAN14 | H in REC |
| CSF2RB | H in REC | IL6R | H in REC | POSTN | H in REC | UBALD2 | H in REC |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| CSTA | H in REC | IL7R | H in REC | PREX1 | H in REC | ULK1 | H in REC |
| CTB-113P19.1 | H in REC | INADL | H in REC | PTGDS | H in REC | UNC119 | H in REC |
| CTTNBP2NL | H in REC | ITGB2 | H in REC | PTPRF | H in REC | UQCR10 | H in REC |
| CWF19L1 | H in REC | ITM2A | H in REC | PVRL1 | H in REC | ZDHHC18 | H in REC |
| CXCL12 | H in REC | KDELR1 | H in REC | PXDN | H in REC | ZFP36 | H in REC |
| CXCL13 | H in REC | KIF21A | H in REC | RASSF6 | H in REC | | |

Genes ≥ 5-fold differentially expressed genes in the recurrence sample compared to the initial DLBL sample 2. (H in REC: High in recurrence), (L in REC: Low in recurrence). Actual data are shown Figure 8b.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| DEC1 | L in REC | DYNLL1 | L in REC | LOC101927502 | L in REC | RAB31 | H in REC |
| ABCB1 | L in REC | EDIL3 | L in REC | LOC101928942 | L in REC | RALYL | L in REC |
| ADARB2 | L in REC | EFHC2 | L in REC | LOC101929239 | L in REC | RBFOX1 | L in REC |
| ADGRB3 | L in REC | ELAVL2 | L in REC | LOC285692 | L in REC | RNF17 | L in REC |
| ADGRF5 | L in REC | EMILIN2 | L in REC | LOC442028 | L in REC | ROS1 | L in REC |
| ADGRL3 | L in REC | FPGT | L in REC | LRP1B | L in REC | SERPINA3 | L in REC |
| AGBL1 | L in REC | FRG2 | L in REC | LSAMP | L in REC | SFTPB | L in REC |
| AGPAT9 | L in REC | GAS2 | L in REC | LTF | L in REC | SLC26A4 | L in REC |
| AK5 | L in REC | GATC | L in REC | LYPD4 | L in REC | SNTG1 | L in REC |
| AMELX | L in REC | GOLGA6L1 | L in REC | MCTP1 | L in REC | SORCS3 | L in REC |
| AMOT | L in REC | GOLGA6L6 | L in REC | MEGF11 | L in REC | SVOPL | L in REC |
| AMPH | L in REC | GOLGA8EP | L in REC | MIR125B2 | L in REC | TKTL1 | L in REC |
| AMY1A | L in REC | GPC4 | L in REC | MIR3612 | L in REC | TMEM167B | L in REC |
| ANKRD30B | L in REC | GPC5 | L in REC | MIR4450 | L in REC | TPTE2 | L in REC |
| APTX | H in REC | GPD1L | L in REC | MIR4789 | L in REC | TRIAP1 | L in REC |
| AS3MT | L in REC | GPR174 | L in REC | MIR6744 | H in REC | USP44 | L in REC |
| BCAP29 | L in REC | GPR85 | L in REC | MIR943 | H in REC | ZAN | L in REC |
| C11orf39 | L in REC | GRM7 | L in REC | MME | H in REC | ZCWPW2 | L in REC |
| C1orf194 | L in REC | HMGA2 | L in REC | MMP1 | H in REC | ZDHHC11 | L in REC |
| CACNA2D3 | L in REC | HTR2C | L in REC | MSI1 | H in REC | ZFP36L2 | H in REC |
| CAMKK1 | L in REC | HYDIN | L in REC | MTHFD2P1 | H in REC | ZPBP | L in REC |
| CELA2A | L in REC | ITGB6 | L in REC | MYOM1 | H in REC | | |
| CELSR2 | L in REC | ITM2A | L in REC | NCAPH | H in REC | | |
| CFI | L in REC | KCNH7 | L in REC | NINL | H in REC | | |
| CNTNAP5 | L in REC | KIAA1324 | L in REC | NKAIN2 | L in REC | | |
| COL11A1 | H in REC | LAMB3 | L in REC | NRG1 | L in REC | | |
| COL12A1 | H in REC | LAMB4 | L in REC | NRXN3 | L in REC | | |
| COQ5 | L in REC | LINC00113 | L in REC | NTN4 | L in REC | | |
| CSK | H in REC | LINC00707 | L in REC | NTRK2 | L in REC | | |
| CSMD1 | L in REC | LINC00824 | L in REC | OPCML | L in REC | | |
| DEPTOR | L in REC | LINC01017 | L in REC | OXSM | L in REC | | |
| DIP2C | L in REC | LINC01603 | L in REC | PAK3 | L in REC | | |
| DPP6 | L in REC | LINC01609 | L in REC | PCSK5 | L in REC | | |
| DSCAM | L in REC | LOC101927082 | L in REC | PDE8B | L in REC | | |

Genes ≥ 5-fold differentially expressed genes in the recurrence sample compared to the initial DLBL sample 3. (H in REC: High in recurrence), (L in REC: Low in recurrence). Actual data are shown Figure 8c.

| | | | | | |
|---|---|---|---|---|---|
| AA06 | L in REC | IL1RAPL1 | L in REC | TRIM29 | H in REC |
| ACY3 | H in REC | IL2RA | L in REC | ZAP70 | L in REC |
| AGBL4 | L in REC | KCNIP4 | L in REC | ZNF804A | L in REC |
| AHNAK2 | H in REC | KLF5 | H in REC | FLG | H in REC |
| APCDD1 | H in REC | KRT1 | H in REC | HLA-DOA | L in REC |
| C4A | L in REC | KRT10 | H in REC | IFI44 | H in REC |
| C7 | L in REC | KRT17 | H in REC | IFI44L | H in REC |
| CARD11 | L in REC | KRT2 | H in REC | SPOCK2 | L in REC |
| CCL21 | L in REC | KRT5 | H in REC | SPP1 | H in REC |
| CD3E | L in REC | KRT6A | H in REC | SPTBN2 | H in REC |
| CD6 | L in REC | LOC100506585 | L in REC | TOX2 | L in REC |
| CDHR1 | H in REC | LOC100507388 | L in REC | | |
| CDT1 | H in REC | LOC101927159 | L in REC | | |
| CMPK2 | H in REC | LRP1B | L in REC | | |
| CNTNAP2 | L in REC | MAGI2 | L in REC | | |
| COL17A1 | H in REC | MIR1203 | L in REC | | |
| COL7A1 | H in REC | MIR1205 | L in REC | | |
| COMP | L in REC | MIR569 | L in REC | | |
| CSMD1 | L in REC | MUC16 | L in REC | | |
| CTNNA3 | L in REC | MX1 | H in REC | | |
| DAB1 | L in REC | NELL2 | L in REC | | |
| DMKN | H in REC | NRXN3 | L in REC | | |
| DSC3 | H in REC | OPCML | L in REC | | |
| DSG1 | H in REC | PARVG | L in REC | | |
| DSP | H in REC | PERP | H in REC | | |
| EPPK1 | H in REC | PIM2 | L in REC | | |
| EVPL | H in REC | PKP1 | H in REC | | |
| F5 | L in REC | PPL | H in REC | | |
| FAT2 | H in REC | PVRL1 | H in REC | | |
| FCER2 | L in REC | RBFOX1 | L in REC | | |
| FCMR | L in REC | SCEL | H in REC | | |
| FDCSP | L in REC | SDC1 | H in REC | | |
| FGF14 | L in REC | SFN | H in REC | | |
| FGFR3 | H in REC | SPINK5 | H in REC | | |

Genes ≥ 5-fold differentially expressed genes in the recurrence sample compared to the initial DLBL sample 4. (H in REC: High in recurrence), (L in REC: Low in recurrence). Actual data are shown Figure 8d.

| | | | |
|---|---|---|---|
| A2M | L in REC | FLNA | L in REC |
| ABI3BP | L in REC | FLT3 | H in REC |
| AK4 | H in REC | FOS | H in REC |
| ALDOC | H in REC | FOSB | H in REC |
| APOD | H in REC | GPNMB | L in REC |
| APOE | L in REC | H1F0 | H in REC |
| APP | L in REC | HLA-DRB5 | L in REC |
| ARRDC3 | H in REC | HLA-DRB6 | L in REC |
| BGN | L in REC | HSPG2 | L in REC |
| BMF | H in REC | IGFBP2 | H in REC |
| C1S | L in REC | JUN | H in REC |
| C3 | L in REC | JUND | H in REC |
| CD68 | L in REC | KIF18B | L in REC |
| CLSPN | L in REC | LOC101929753 | H in REC |
| COL12A1 | L in REC | LYZ | L in REC |
| COL15A1 | L in REC | MMP2 | L in REC |
| COL1A1 | L in REC | MXRA5 | L in REC |
| COL1A2 | L in REC | NR4A1 | H in REC |
| COL3A1 | L in REC | NXPH4 | H in REC |
| COL4A1 | L in REC | PMAIP1 | H in REC |
| COL4A2 | L in REC | PNRC1 | H in REC |
| COL5A2 | L in REC | POSTN | L in REC |
| COL6A1 | L in REC | PPP1R15A | H in REC |
| COL6A2 | L in REC | REXO2 | H in REC |
| COL6A3 | L in REC | RGS1 | H in REC |
| CTB-113P19.1 | L in REC | SESN1 | H in REC |
| CXCL9 | L in REC | SULF1 | L in REC |
| DCN | L in REC | SYN3 | L in REC |
| DUSP1 | H in REC | TAGAP | H in REC |
| E2F2 | L in REC | THBS1 | L in REC |
| EGR1 | H in REC | TMEM132B | H in REC |
| EPAS1 | L in REC | TNC | L in REC |
| F13A1 | L in REC | VEGFA | H in REC |
| FBN1 | L in REC | VWF | L in REC |
| | | ZBTB16 | H in REC |

Genes ≥ 5-fold differentially expressed genes in the recurrence sample compared to the initial DLBL sample 5. (H in REC: High in recurrence), (L in REC: Low in recurrence). Actual data are shown Figure 8e.

| | | | |
|---|---|---|---|
| ADD2 | H in REC | HSD11B1 | L in REC |
| ALDH1A1 | L in REC | HTR3A | L in REC |
| ALOX15B | L in REC | HTRA4 | L in REC |
| AMICA1 | L in REC | JCHAIN | H in REC |
| ANKRD22 | L in REC | MIR3161 | L in REC |
| ATRNL1 | H in REC | MIR3664 | H in REC |
| BCAS1 | H in REC | MIR511 | L in REC |
| C2 | L in REC | MMP20 | L in REC |
| C3AR1 | L in REC | MT1G | L in REC |
| CAPNS2 | L in REC | MUC4 | H in REC |
| CCL18 | L in REC | MZB1 | H in REC |
| CCL19 | L in REC | NXT2 | H in REC |
| CCRL2 | L in REC | ORM2 | L in REC |
| CD209 | L in REC | PADI2 | L in REC |
| CD6 | L in REC | PEG10 | H in REC |
| CHI3L1 | L in REC | PLEKHG3 | L in REC |
| CHIT1 | L in REC | PMEL | L in REC |
| CLEC12A | L in REC | POSTN | H in REC |
| COL12A1 | H in REC | PSTPIP2 | L in REC |
| CPM | L in REC | RYR1 | L in REC |
| CTAGE9 | H in REC | S100A8 | L in REC |
| CTGF | H in REC | S100A9 | L in REC |
| CTNND2 | L in REC | SCHLAP1 | L in REC |
| CYP1B1 | L in REC | SERPINA1 | L in REC |
| CYR61 | H in REC | SIGLEC1 | L in REC |
| DPCR1 | L in REC | SIGLEC14 | L in REC |
| DUSP5P1 | L in REC | SRC | L in REC |
| FAM171B | H in REC | SULF1 | H in REC |
| FMN1 | L in REC | THBS1 | H in REC |
| FOXP2 | H in REC | TLR8 | L in REC |
| GPR160 | H in REC | TNFSF13B | L in REC |
| GPX3 | L in REC | HLA-DQA2 | L in REC |

Genes ≥ 5-fold differentially expressed genes in the recurrence sample compared to the initial DLBL sample 6. (H in REC: High in recurrence), (L in REC: Low in recurrence). Actual data are shown Figure 8f.

| | |
|---|---|
| CAMK4 | H in REC |
| COL15A1 | L in REC |
| COLEC12 | H in REC |
| FARP1 | H in REC |
| GPR174 | H in REC |
| HTR3A | H in REC |
| LINGO1 | H in REC |
| MMP9 | H in REC |
| NRCAM | H in REC |
| POF1B | H in REC |
| SLC26A3 | H in REC |
| SMARCA1 | H in REC |

Genes ≥ 5-fold differentially expressed genes in the recurrence sample compared to the initial DLBL sample 7. (H in REC: High in recurrence), (L in REC: Low in recurrence). Actual data are shown Figure 8g.

| | |
|---|---|
| ATRNL1 | H in REC |
| CHIT1 | L in REC |
| CUEDC1 | H in REC |
| DCC | L in REC |
| JCHAIN | L in REC |
| MGMT | L in REC |
| MS4A1 | L in REC |
| POSTN | H in REC |
| SVEP1 | L in REC |